(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

POWELL MURCHISON
Managing Editor

HELEN M. DAVIS
Associate Editor

Editorial Board

GORDON ALLPORT
Harvard University

JOHN A. ANTOINETTI
Life Insurance Agency
Management Association

ARTHUR G. BILLS
University of Cincinnati

E. A. BOTT
University of Toronto

LEONARD CARMICHAEL
National Geographic Society

HERBERT S. CONRAD
U. S. Office of Education

JOHN F. DASHIELL
Florida State University

LEONARD W. DOOB
Yale University

ROY M. DORCUS
University of California
at Los Angeles

FRANK A. GELDARD
Princeton University

CLARENCE H. GRAHAM
Columbia University

ERNEST R. HILGARD
Stanford University

JOHN E. HORROCKS
Ohio State University

HEINRICH KLÜVER
University of Chicago

NORMAN L. MUNN
University of Adelaide,
Australia

JOHN PAUL NAFE
Florida State University

WARREN H. TEICHNER
Northeastern University

LEE EDWARD TRAVIS
University of Southern
California

ROBERT C. TRYON
University of California

JOSEPH ZUBIN
Psychiatric Institute,
Columbia University

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 68

1968

Copyright © 1968 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved

Published Bimonthly at 35 New St., Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts

AUTHOR INDEX

Alexander, Theron 215

Ball, Richard J. 55

Baron, Robert A. 141

Bartley, S. Howard 55

Bell, Graham B. 49

Berry, Newell H. 63

Botwinick, Jack 167

Brandt, Lewis W. 261

Braun, John R. 3

Breglio, Vincent J. 291

Coyle, F. A., Jr. 305

Cutter, Henry S. G. 21

Davidson, P. O. 249

DeLuca, Joseph 83, 145

Drakeford, Grant C. 223

Dreistadt, Roy 97

Finney, Joseph C. 321

Flanagan, Bruce C. 185

Fowler, Raymond D., Jr. 305

Friedman, S. Thomas 281

Gilbert, James E. 173

Gilbert, Joseph 173

Gladstone, Roy 33

Goss, Allen 287

Heilizer, Frederick H. 21

Herring, Fred H. 211

Horn, John 151

Horne, E. P. 185

Hunter, E. C. 193

Juliard, Alex S. 267

Juliard, Andre L. 267

Keeran, Charles V., Jr. 49

Kingston, Albert J., Jr. 87

Kirk, Charles 215

LaFaro, Dolores 3

LaGaipa, John J. 257

Lane, Ellen A. 15

Lester, David 243

Lester, Gene 9

Lind, Dennis 151

Mahrer, Alvin 151

Miller, Mercedes 33

Morosko, Tom 287

Munz, David C. 181

Nelson, Paul D. 63

Oram, Phyllis G. 39

Pedersen, Darhl M. 291

Pustel, Gabriel 299

Quereshi, Mohammed Y. 73

Reed, P. C. 305

Resnikoff, Arthur 117

Rice, James A. 159

Roberts, William A. 227

Saltzstein, Herbert D. 313

Schauble, Paul G. 117

Schrag, A. R. 249

Sheehan, Peter W. 203

Sheldon, Robert 287

Sherrill, David 281

Siegel, Louis 299

Smouse, Albert D. 181

Snider, James G. 223

Sternlicht, Manny 299

Stewart, Patricia 151

Stoyle, Judith 215

Thompson, Larry W. 167

Voevodsky, John 129

Weaver, Wendell W. 87

Wertheimer, Michael 211

Weyant, Robert G. 121

White, Chester A. 305

White, William F. 87

Woody, Robert H. 117

Yin, Robert K. 313

TABLE OF CONTENTS

Fakability of the Sixteen Personality Factor Questionnaire, Form C . . . 3
By John R. Braun and Dolores LaFaro

The case for efferent change during prism adaptation 9
By Gene Lester

The influence of sex and race on process-reactive ratings of schizophrenics . 15
By Ellen A. Lane

The development of some procedures for the assessment of risk 21
By Henry S. G. Cutter and Frederick H. Heilizer

Rationality, motivation, and extinction 33
By Roy Gladstone and Mercedes Miller

Induction of action and attitude change: The function of role-self conflicts and levels of endorsement 39
By Phyllis G. Oram

Reading ease as a factor in improved communication effectiveness . . . 49
By Charles V. Keeran, Jr. and Graham B. Bell

Brightness and saturation changes resultant from temporal manipulation of spectral photic input for normal and color-deficient observers . . . 55
By Richard J. Ball and S. Howard Bartley

Cohesion in Marine recruit platoons 63
By LCDR Paul D. Nelson, MSC, USN, and CDR Newell H. Berry, MSC, USN

The comparability of WAIS and WISC subtest scores and *IQ* estimates . . 73
By Mohammed Y. Quereshi

Predicting the full scale WAIS *IQ* of Army basic trainees 83
By Joseph DeLuca

Affective correlates of reading comprehension 87
By Wendell W. Weaver, William F. White, and Albert J. Kingston, Jr.

An analysis of the use of analogies and metaphors in science 97
By Roy Dreistadt

Personality correlates of withdrawal from smoking 117
By Arthur Resnikoff, Paul G. Schauble, and Robert H. Woody

Relevant and irrelevant cues in spatial reversal learning 121
By Robert G. Weyant

A mathematical representation of time-variant quantitative behavior . . 129
By John Voevodsky

Authoritarianism, locus of control, and risk taking 141
By Robert A. Baron

Psychosexual conflict in adolescent enuretics 145
By Joseph DeLuca

Symptom patterns in psychiatric patients: A goal-directed approach to psychiatric symptomatology 151
By Alvin Mahrer, Patricia Stewart, John Horn, and Dennis Lind

Form equivalence of the revised Visual Retention Test in mental retardates: An extension 159
By James A. Rice

Age differences in the relationship between EEG arousal and reaction time . 167
By Larry W. Thompson and Jack Botwinick

Rater reliability of a depressive rating scale 173
By Joseph Gilbert and James E. Gilbert

The effects of anxiety and item difficulty sequence on achievement testing scores 181
By Albert D. Smouse and David C. Munz

Short term memory and coding behavior as a function of set during sequential trials 185
By E. P. Horne and Bruce C. Flanagan

Results of the National Teacher Examinations at Tulane University, 1948-1966 . 193
By E. C. Hunter

Color response to the TAT: An instance of eidetic imagery? 203
By Peter W. Sheehan

Individual differences in figural aftereffects: Some problems and potentials . 211
By Michael Wertheimer and Fred H. Herring

The language of children in the "inner city" 215
By Theron Alexander, Judith Stoyle, and Charles Kirk

Two scales to measure cognitive rigidity 223
By James G. Snider and Grant C. Drakeford

An analysis of multitrial free recall learning with input-order held constant . 227
By William A. Roberts

Attempted suicide as a hostile act 243
By David Lester

The role of the father in guidance clinic consultations 249
By P. O. Davidson and A. R. Schrag

Programmed instruction, teacher ability, and subject matter difficulty . . . 257
By John J. La Gaipa

A note on the use of judges 261
By Lewis W. Brandt

A psychological analysis of "resistance" fighters during World War II . . 267
By Andre L. Juliard and Alex S. Juliard

Interpersonal attraction as a function of freedom of choice 281
By S. Thomas Friedman and David Sherrill

Use of the Edwards Personal Preference Schedule with alcoholics in a vocational rehabilitation program 287
By Allen Goss, Tom Morosko, and Robert Sheldon

The correlation of two self-disclosure inventories with actual self-disclosure: A validity study 291
By Darhl M. Pedersen and Vincent J. Breglio

Institutionalized retardates' reactions to authority figures: God and monster drawings 299
By Gabriel Pustel, Manny Sternlicht, and Louis Siegel

The MMPI measurement of phenotypic anxiety: I. Exploratory study . . . 305
By Raymond D. Fowler, Jr., F. A. Coyle, Jr., P. C. Reed, and Chester A. White

Transfer of social influence effects on psychophysical judgments 313
By Robert K. Yin and Herbert D. Saltzstein

Judgments of ethnic groups 321
By Joseph C. Finney

Age differences in the relationship between EEG arousal and reaction time . 167
By Larry W. Thompson and Jack Botwinick

Rater reliability of a depressive rating scale 173
By Joseph Gilbert and James E. Gilbert

The effects of anxiety and item difficulty sequence on achievement testing scores 181
By Albert D. Smouse and David C. Munz

Short term memory and coding behavior as a function of set during sequential trials 185
By E. P. Horne and Bruce C. Flanagan

Results of the National Teacher Examinations at Tulane University, 1948-1966 . 193
By E. C. Hunter

Color response to the TAT: An instance of eidetic imagery? 203
By Peter W. Sheehan

Individual differences in figural aftereffects: Some problems and potentials . 211
By Michael Wertheimer and Fred H. Herring

The language of children in the "inner city" 215
By Theron Alexander, Judith Stoyle, and Charles Kirk

Two scales to measure cognitive rigidity 223
By James G. Snider and Grant C. Drakeford

An analysis of multitrial free recall learning with input-order held constant . 227
By William A. Roberts

Attempted suicide as a hostile act 243
By David Lester

The role of the father in guidance clinic consultations 249
By P. O. Davidson and A. R. Schrag

Programmed instruction, teacher ability, and subject matter difficulty . . . 257
By John J. La Gaipa

A note on the use of judges 261
By Lewis W. Brandt

A psychological analysis of "resistance" fighters during World War II . . 267
By Andre L. Juliard and Alex S. Juliard

Interpersonal attraction as a function of freedom of choice 281
By S. Thomas Friedman and David Sherrill

Use of the Edwards Personal Preference Schedule with alcoholics in a vocational rehabilitation program 287
By Allen Goss, Tom Morosko, and Robert Sheldon

The correlation of two self-disclosure inventories with actual self-disclosure: A validity study 291
By Darhl M. Pedersen and Vincent J. Breglio

Institutionalized retardates' reactions to authority figures: God and monster drawings 299
By Gabriel Pustel, Manny Sternlicht, and Louis Siegel

The MMPI measurement of phenotypic anxiety: I. Exploratory study . . . 305
By Raymond D. Fowler, Jr., F. A. Coyle, Jr., P. C. Reed, and Chester A. White

Transfer of social influence effects on psychophysical judgments 313
By Robert K. Yin and Herbert D. Saltzstein

Judgments of ethnic groups 321
By Joseph C. Finney

Journal of Genetic Psychology

Founded by G. Stanley Hall in 1891; reorganized by Carl Murchison in 1925

Quarterly, published in March, June, September, and December. Devoted to child behavior, animal behavior, and the problems of aging. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1891 at $15.00 per volume, plus transportation.

Journal of General Psychology

Founded in 1927 by E. B. Titchener and Carl Murchison

Quarterly, published in January, April, July, and October. Devoted primarily to experimental, theoretical, clinical, and historical psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1927 at $15.00 per volume, plus transportation.

Journal of Social Psychology

Founded in 1929 by John Dewey and Carl Murchison

Bimonthly, beginning with February. National, group, cultural, racial, and differential psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Bimonthly, three volumes per year; per annum, $30.00; per volume, $15.00; single numbers, $7.50. First number is dated February, 1930. Complete sets from 1930 at $15.00 per volume, plus transportation.

Each author will receive 100 unbound reprints gratis. Additional reprints may be purchased at a nominal cost when the author returns page-proof.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Journal of Genetic Psychology

Founded by G. Stanley Hall in 1891; reorganized by Carl Murchison in 1925

Quarterly, published in March, June, September, and December. Devoted to child behavior, animal behavior, and the problems of aging. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1891 at $15.00 per volume, plus transportation.

Journal of General Psychology

Founded in 1927 by E. B. Titchener and Carl Murchison

Quarterly, published in January, April, July, and October. Devoted primarily to experimental, theoretical, clinical, and historical psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1927 at $15.00 per volume, plus transportation.

Journal of Social Psychology

Founded in 1929 by John Dewey and Carl Murchison

Bimonthly, beginning with February. National, group, cultural, racial, and differential psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Bimonthly, three volumes per year; per annum, $30.00; per volume, $15.00; single numbers, $7.50. First number is dated February, 1930. Complete sets from 1930 at $15.00 per volume, plus transportation.

Each author will receive 100 unbound reprints gratis. Additional reprints may be purchased at a nominal cost when the author returns page-proof.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Journal of Genetic Psychology

Founded by G. Stanley Hall in 1891; reorganized by Carl Murchison in 1925

Quarterly, published in March, June, September, and December. Devoted to child behavior, animal behavior, and the problems of aging. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1891 at $15.00 per volume, plus transportation.

Journal of General Psychology

Founded in 1927 by E. B. Titchener and Carl Murchison

Quarterly, published in January, April, July, and October. Devoted primarily to experimental, theoretical, clinical, and historical psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1927 at $15.00 per volume, plus transportation.

Journal of Social Psychology

Founded in 1929 by John Dewey and Carl Murchison

Bimonthly, beginning with February. National, group, cultural, racial, and differential psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Bimonthly, three volumes per year; per annum, $30.00; per volume, $15.00; single numbers, $7.50. First number is dated February, 1930. Complete sets from 1930 at $15.00 per volume, plus transportation.

Each author will receive 100 unbound reprints gratis. Additional reprints may be purchased at a nominal cost when the author returns page-proof.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Journal of Genetic Psychology

Founded by G. Stanley Hall in 1891; reorganized by Carl Murchison in 1925

Quarterly, published in March, June, September, and December. Devoted to child behavior, animal behavior, and the problems of aging. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1891 at $15.00 per volume, plus transportation.

Journal of General Psychology

Founded in 1927 by E. B. Titchener and Carl Murchison

Quarterly, published in January, April, July, and October. Devoted primarily to experimental, theoretical, clinical, and historical psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1927 at $15.00 per volume, plus transportation.

Journal of Social Psychology

Founded in 1929 by John Dewey and Carl Murchison

Bimonthly, beginning with February. National, group, cultural, racial, and differential psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Bimonthly, three volumes per year; per annum, $30.00; per volume, $15.00; single numbers, $7.50. First number is dated February, 1930. Complete sets from 1930 at $15.00 per volume, plus transportation.

Each author will receive 100 unbound reprints gratis. Additional reprints may be purchased at a nominal cost when the author returns page-proof.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Club Rate for Five International Psychological Journals

	PER ANNUM
THE JOURNAL OF GENETIC PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
GENETIC PSYCHOLOGY MONOGRAPHS (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF GENERAL PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF SOCIAL PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
THE JOURNAL OF PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
TOTAL	$120.00
TEACHERS AND STUDENTS EDITION, *Restricted Sale*	$50.00

The *Teachers and Students Edition* is open to any individual psychologist, teacher, student, or physician but requires an agreement that the edition will not be offered for sale till after five years from publication and an agreement that the edition will not be substituted for an institutional subscription.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Genetic Psychology Monographs

Founded in 1925 by Carl Murchison

Quarterly: February, May, August, November. Two volumes per year. Child behavior, animal behavior, and the problems of aging. Each number one or more complete researches. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1926 at $15.00 per volume, plus transportation.

NOTICE. Effective immediately we offer prior publication for certain selected monographs. Charges will be based on the announced rates for *The Journal of Psychology. For a limited period of time our policy is changed for regular monographs: Authors will be charged for cuts and expensive tables, but no page charge will be imposed.* Two hundred reprints of each monograph will be supplied gratis.

The listing of monographs for 1926 through 1963 will be printed in the first half of each volume of all five journals. Monographs published after 1963 will be listed in the second half of each volume.

Genetic Psychology Monographs *(continued)*

VOLUME 1—1926

1. Performance tests for children of pre-school age—R. Stutsman
2. An experimental study of the eidetic type—H. Klüver

3 & 4. A study of natio-racial mental differences—N. D. M. Hirsch

5. A psychological study of juvenile delinquency by group methods—J. W. Bridges and K. M. B. Bridges
6. The influence of puberty praecox upon mental growth—A. Gesell

VOLUME 2—1927

1 & 2. The mind of a gorilla—R. M. Yerkes

3. The role of eye-muscles and mouth-muscles in the expression of the emotions—K. Dunlap
4. Family similarities in mental-test abilities—R. R. Willoughby
5. Coordination in the locomotion of infants—L. H. Burnside
6. The mind of a gorilla: Part II. Mental development—R. M. Yerkes

VOLUME 3—January-June, 1928

1. An experimental study of the olfactory sensitivity of the white rat—J. R. Liggett
2. A photographic study of eye movements in reading formulae—M. A. Tinker
3. An experimental study of the East Kentucky mountaineers—N. D. M. Hirsch
4. Responses of foetal guinea pigs prematurely delivered—G. T. Avery
5. Objective differentiation between three groups in education (teachers, research workers, and administrators)—M. B. Jensen
6. The effect of segregation on the sex behavior of the white rat as measured by the obstruction method—M. Jenkins

VOLUME 4—July-December, 1928

1. Observation and training of fundamental habits in young children—E. A. Bott, W. E. Blatz, N. Chant, and H. Bott

2 & 3. Determination of a content of the course in literature of a suitable difficulty for junior and senior high school students—M. C. Burch

4 & 5. Methods for diagnosis and treatment of cases of reading disability—M. Monroe

6. The relative effectiveness of lecture and individual reading as methods of college teaching—E. B. Greene

VOLUME 5—January-June, 1929

1. The age factor in animal learning: I. Rats in the problem box and the maze—C. P. Stone
2. The effect of delayed incentive on the hunger drive in the white rat—E. L. Hamilton
3. Which hand is the eye of the blind?—J. M. Smith
4. The effect of attitude on free word association-time—A. G. Ekdahl
5. The localization of tactual space: A study of average and constant errors under different types of localization—L. E. Cole
6. The effects of gonadectomy, vasotomy, and injections of placental and orchic extracts on the sex behavior of the white rat—H. W. Nissen

VOLUME 6—July-December, 1929

1. Learning and growth in identical infant twins: An experimental study by the method of co-twin control—A. Gesell and H. Thompson
2. The age factor in animal learning: II. Rats on a multiple light discrimination box and a difficult maze—C. P. Stone
3. The acquisition and interference of motor habits in young children—E. McGinnis
4. A vocational and socio-educational survey of graduates and non-graduates of small high schools of New England—A. D. Mueller

5 & 6. A study of the smiling and laughing of infants in the first year of life—R. W. Washburn

VOLUME 7—January-June, 1930

1. Tensions and emotional factors in reaction—E. Duffy
2. Teacher influence on class achievement: A study of the relationship of estimated teaching ability to pupil achievement in reading and arithmetic—H. R. Taylor

3 & 4. A study of the effect of inverted retinal stimulation upon spatially coordinated behavior—P. H. Ewert

5. A study of the mental development of children with lesion in the central nervous system—E. E. Lord
6. An experimental study upon three hundred school children over a six-year period—N. D. M. Hirsch

VOLUME 8—July-December, 1930

1. The amount and nature of activities of newborn infants under constant external stimulating conditions during the first ten days of life—O. C. Irwin
2. Race and social differences in performance tests—S. D. Porteus, *et al.*
3. Language and growth: The relative efficacy of early and deferred vocabulary training, studied by the method of co-twin control—L. C. Strayer
4. Eye-movements and optic nystagmus in early infancy—J. M. McGinnis

5 & 6. Reactions of kindergarten, first-, and second-grade children to constructive play materials—L. Farwell

VOLUME 9—January-June, 1931

1 & 2. The status of the first-born with special reference to intelligence—H. H. Hsiao

3 & 4. An experimental study of bright, average, and dull children at the four-year mental level—H. P. Davidson

5. An historical, critical, and experimental study of the Seashore-Kwalwasser test battery—P. R. Farnsworth
6. A comparison of difficulty and improvement in the learning of bright and dull children in reproducing a descriptive selection—F. T. Wilson

VOLUME 10—July-December, 1931

1. A comparative study of a group of southern white and negro infants—M. B. McGraw

2 & 3. An experimental study of prehension in infants by means of systematic cinema records—H. M. Halverson

4. The limits of learning ability in kittens—A. M. Shuey

5 & 6. The effect of habit interference upon performance in maze learning—O. W. Alm

VOLUME 11—January-June, 1932

1. General factors in transfer of training in the white rat—T. A. Jackson
2. The effect of color on visual apprehension and perception—M. A. Tinker
3. The reliability and validity of maze experiments with white rats—R. Leeper
4. A critical study of two lists of best books for children—F. K. Shuttleworth

5 & 6. Measuring human energy cost in industry: A general guide to the literature—R. M. Page

VOLUME 12—July-December, 1932

1. Family resemblances in verbal and numerical abilities—H. D. Carter
2. The development of fine prehension in infancy—B. M. Castner

3 & 4. The growth of adaptive behavior in infants: An experimental study at seven age levels—H. M. Richardson

5 & 6. Differential reactions to taste and temperature stimuli in newborn infants—K. Jensen

VOLUME 13—January-June, 1933

1. A critique of sublimation in males: A study of forty superior single men—W. S. Taylor
2. A study of the nature, measurement, and determination of hand preference—H. L. Koch, *et al.*
3. The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty—H. E. Jones and H. S. Conrad
4. The relation between the complexity of the habit to be acquired and the form of the learning curve in young children—M. L. Mattson
5. Eating habits in relation to personality development of two- and three-year-old children: A study of sixty-nine children in two nursery schools—A. A. Eliot
6. Coordinating mechanisms of the spinal cord—O. C. Ingebritsen

Genetic Psychology Monographs *(continued)*

VOLUME 14—July-December, 1933

1. Mental growth during the first three years: A developmental study of sixty-one children by repeated tests—N. Bayley
2. A study of triplets: including theories of their possible genetic relationships—F. N. Anderson and N. V. Scheidemann
3. The objective measurement of emotional reactions—H. V. Gaskill
4. Development of behavior in the fetal cat—J. D. Coronios
5. A study of certain language developments of children in grades four to twelve, inclusive—L. L. LaBrant
6. The effect of early and delayed practice on memory and motor performances studied by the method of co-twin control—J. R. Hilgard

VOLUME 15—January-June, 1934

1. Studies in the psychology of tone and music—P. R. Farnsworth
2. Motor learning of children in equilibrium in relation to nutrition—E. L. Beebe
3. Discrimination limens of pattern and size in the goldfish *Carassius auratus*—J. B. Rowley
4. Limits of learning ability in the white rat and the guinea pig—B. F. Riess

5 & 6. The limits of learning ability in rhesus monkeys—H. A. Fjeld

VOLUME 16—July-December, 1934

1. A statistical study of ratings on the California Behavior Inventory for Nursery-School Children—H. S. Conrad
2. An eye-movement study of objective examination questions—A. Frandsen
3. An experimental study of constitutional types—O. Klineberg, S. E. Asch, and H. Block
4. The development of a battery of objective group tests of manual laterality, with the results of their application to 1300 children—W. N. Durost

5 & 6. An experimental study in the prenatal guinea-pig of the origin and development of reflexes and patterns of behavior in relation to the stimulation of specific receptor areas during the period of active fetal life—L. Carmichael

VOLUME 17—January-December, 1935

1. Organization of behavior in the albino rat—R. L. Thorndike
2. Brightness discrimination in the rhesus monkey—M. P. Crawford
3. The limits of learning ability in cebus monkeys—A. M. Koch
4. Nature-nurture and intelligence—A. M. Leahy
5. On intelligence of epileptic children—E. B. Sullivan and L. Gahagan
6. A study of the play of children of preschool age by an unobserved observer—D. L. Cockrell

VOLUME 18—January-December, 1936

1. Sex differences in variational tendency—Q. McNemar and L. M. Terman
2. The process of learning to dress among nursery-school children—C. B. Key, M. R. White, M. P. Honzik, A. B. Heiney, and D. Erwin
3. A study of the present social status of a group of adults, who, when they were in elementary schools, were classified as mentally deficient—W. R. Baller
4. The influence of specific experience upon mental organizations—A. Anastasi

5 & 6. Studies in aggressiveness—L. Bender, S. Keiser, and P. Schilder

VOLUME 19—January-December, 1937

1. Psychological bases of self-mutilation—C. Dabrowski
   Masculine temperament and secondary sex characteristics: A study of the relationship between psychological and physical measures of masculinity—H. Gilkinson
2. A psychological study of forty unmarried mothers—R. D. Nottingham
   Behavior problems in the children of psychotic and criminal parents—L. Bender
3. Domination and integration in the social behavior of young children in an experimental play situation—H. H. Anderson
4. The sequential patterning of prone progression in the human infant—L. B. Ames

VOLUME 20—January-December, 1938

1. The relationship between characteristics of personality and physique in adolescents—P. S. de Q. Cabot
2. Behavior problems of elementary school children: A descriptive and comparative study—I. Y. Masten
   Graphic representation of a man by four-year-old children in nine prescribed drawing situations—P. F. Gridley
3. Differences between two groups of adult criminals—R. S. Tolman
4. A comparative study by means of the Rorschach method of personality development in twenty pairs of identical twins—E. Troup
   Individual differences in the facial expressive behavior of preschool children: A study by the time-sampling method—C. Swan

VOLUME 21—January-December, 1939

1. An experimental analysis of "level of aspiration"—R. Gould
2. Some light on the problem of bilingualism as found from a study of the progress in mastery of English among preschool children of non-American ancestry in Hawaii—M. E. Smith
3. Domination and social integration in the behavior of kindergarten children and teachers—H. H. Anderson
   The capacity of the rhesus and cebus monkey and the gibbon to acquire differential response to complex visual stimuli—W. E. Galt
4. The social-sex development of children—E. H. Campbell

VOLUME 22—January-December, 1940

1. Measuring human relations: An introduction to the study of the interaction of individuals—E. D. Chapple
2. Aggressive behavior in young children and children's attitudes toward aggression—M. D. Fite
3. Student attitudes toward religion—E. Nelson
   The prediction of the outcome-on-furlough of dementia praecox patients—J. S. Jacob
   Significant characteristics of preschool children as located in the Conrad inventory—K. H. Read
4. Learning by children at noon-meal in a nursery school: Ten "good" eaters and ten "poor" eaters—J. B. McCay, E. B. Waring, and P. J. Kruse
   Studies in the interpretation of play: I. Clinical observation of play disruption in young children—E. H. Erikson

VOLUME 23—January-June, 1941

1. An analysis of certain variables in a developmental study of language—F. M. Young
   Infant development under conditions of restricted practice and of minimum social stimulation—W. Dennis
   An analysis of the mental factors of various age groups from nine to sixty—B. Balinsky
2. Factors influencing performance on group and individual tests of intelligence: I. Rate of work—M. W. Bennett
   Individual differences in apperceptive reaction: A study of the response of preschool children to pictures—E. W. Amen

VOLUME 24—July-December, 1941

1. Twins *T* and *C* from infancy to adolescence: A biogenetic study of individual differences by the method of co-twin control—A. Gesell and H. Thompson
   Finger nail-biting: Its incipiency, incidence, and amelioration—A. L. Billig
   An experimental study of the factors of maturation and practice in the behavioral development of the embryo of the frog, *Rana pipiens*—A. Fromme
2. The Fels child behavior scales—T. W. Richards and M. P. Simons
   Measurement of the size of general English vocabulary through the elementary grades and high school—M. K. Smith
   Stereotypes in the field of musical eminence—P. R. Farnsworth

VOLUME 25—January-June, 1942

1. A study of factors determining family size in a selected professional group—J. C. Flanagan
   A genetic study of geometrical-optical illusions—A. Walters
   Interpretation of behavior-ratings in terms of favorable and unfavorable deviations: A study of scores from the Read-Conrad Behavior Inventory—K. H. Read and H. S. Conrad
2. Are there any innate behavior tendencies?—J. B. Schoolland
   An investigation of the intelligibility of the speech of the deaf—C. V. Hudgins and F. C. Numbers

Genetic Psychology Monographs *(continued)*

VOLUME 26—July-December, 1942

1. The critical frequency limen for visual flicker in children between the ages of 6 and 18—V. L. Miller
   Some factors determining handedness in the white rat—K. L. Wentworth
2. Motivation and behavior—E. Frenkel-Brunswik

VOLUME 27—January-June, 1943

1. Comparison of children's personality traits, attitudes, and intelligence with parental occupation—N. R. Maddy
2. A comparative study of mental functioning patterns of problem and non-problem children seven, eight, and nine years of age—M. L. Pignatelli

VOLUME 28—July-December, 1943

1. Separation anxiety in young children: A study of hospital cases—H. Edelston
2. Correlates of vocational preferences—W. A. Bradley, Jr.

VOLUME 29—January-June, 1944

1. Mental changes after bilateral prefrontal lobotomy—S. D. Porteus and R. D. Kepner
2. A twin-controlled experiment on the learning of auxiliary languages—B. Price, W. J. Kostir, and W. M. Taylor

VOLUME 30—July-December, 1944

1. A method of administering and evaluating the thematic appreciation test in group situations—R. M. Clark
2. A study of anxiety reactions in young children by means of a projective technique—R. Temple and E. W. Amen

VOLUME 31—January-June, 1945

1. The evolution of intelligent behavior in rhesus monkeys—B. Weinstein
2. Perceptual behavior of brain-injured, mentally defective children: An experimental study by means of the Rorschach technique—H. Werner

VOLUME 32—July-December, 1945

1. A clinical study of sentiments: I.—H. A. Murray and C. D. Morgan
2. A clinical study of sentiments: II.—H. A. Murray and C. D. Morgan

VOLUME 33—January-June, 1946

1. Interpretation of spontaneous drawings and paintings—T. S. Waehner
   Preferences for sex symbols and their personality correlates—K. Franck
2. Outstanding traits: In a selected college group, with some reference to career interests and war records—F. L. Wells and W. L. Woods

VOLUME 34—July-December, 1946

1. The relation of emotional adjustment to intellectual function—J. L. Despert and H. O. Pierce
   The smiling response: A contribution to the ontogenesis of social relations—R. A. Spitz
2. Finger-painting and personality diagnosis—P. J. Napoli

VOLUME 35—January-June, 1947

1. The thematic apperception technique in the study of culture-personality relations—W. E. Henry
2. A continuation study of anxiety reactions in young children by means of a projective technique—M. Dorkey and E. W. Amen
   A study of the vocational interest trends of secondary school and college women—A. M. Cawley

VOLUME 36—July-December, 1947

1. Maze test validation and psychosurgery—S. D. Porteus and H. N. Peters
2. The diagnostic implications of Rorschach's test in case studies of mental defectives—I. Jolles

VOLUME 37—January-June, 1948

1. The radio day time serial: A symbolic analysis—W. L. Warner and W. E. Henry
   The relation of personality characteristics and response to verbal approval in a learning task—G. L. Grace
2. The mechanism of vision: XVIII. Effects of destroying the visual "associative areas" of the monkey—K. S. Lashley
   A study of the relationship between handwriting and personality variables—P. Castelnuova-Tedesco

VOLUME 38—July-December, 1948

1. Modern language learning: The intensive course as sponsored by the United States Army, and implications for the undergraduate course of study—M. Lind
   Conflict: A study of some interactions between appetite and aversion in the white rat—M. A. Tolcott
2. Schizophrenia and the *MAPS* test: A study of certain formal psycho-social aspects of fantasy production in schizophrenia as revealed by performance on the Make a Picture Story (*MAPS*) Test—E. S. Shneidman
   A study of the transmission of authority patterns in the family—H. L. Ingersoll

VOLUME 39—January-June, 1949

1. A study of the psychoanalytic theory of psychosexual development—G. S. Blum
   The assessment of parental attitudes in relation to child adjustment—E. J. Shoben, Jr.
2. Qualitative differences in the vocabulary responses of normals and abnormals—H. Feifel
   The relative effectiveness of motion and still pictures as stimuli for eliciting fantasy stories about adolescent-parent relationships—P. E. Eiserer
   The organization of hereditary maze-brightness and maze-dullness—L. V. Searle

VOLUME 40—July-December, 1949

1. An experimental study of what young school children expect from their teachers—B. Biber and C. Lewis
   A study of the relative effects of age and of test difficulty upon factor patterns—H. A. Curtis
   A projective experiment using incomplete stories with multiple choice endings—J. K. Seaton
2. Effects of sex role and social status on the early adolescent personality—E. Milner
   Social perceptions and attitudes of children—M. Radke, H. Trager, and H. Davis

VOLUME 41—January-June, 1950

1. Some psychological and educational aspects of pediatric practice: A study of well-baby clinics—L. H. Blum
   One-trial learning in the domestic rat—B. B. Hudson
   An introduction to the principles of scientific psychoanalysis—A. Ellis
2. Awareness of racial differences by preschool children in Hawaii—D. V. Springer
   Age trends in children's evaluation of teacher-approved and teacher-disapproved behavior—S. L. Witryol
   The relationship between level of vocational aspiration and certain personal data: A study of some traits and influences bearing on the prestige level of vocational choice—J. Stubbins

VOLUME 42—July-December, 1950

1. Personality patterns of suicidal mental hospital patients—N. L. Farberow
   Sex-rôle identification in young children in two diverse social groups—M. Rabban
2. A study of the influence of the social field on individual behavior: As revealed in the expression of hostility and warmth by neurotics and paranoid schizophrenics in discussion group situations—D. Shapiro
   An experimental study of avoidance—R. F. Hefferline

VOLUME 43—January-June, 1951

1. A study of copying ability in children—E. A. Townsend
   Prestige motivation of gifted children—D. P. Ausubel
2. A psychological study of physical scientists—A. Roe

VOLUME 44—July-December, 1951

1. The organization of hostility controls in various personality structures—S. Fisher and E. Hinds
   Children and radio: A study of listeners and non-listeners to various types of radio programs in terms of selected ability, attitude, and behavior measures—E. A. Ricciuti
2. Quantitative expression in young children—W. E. Martin
   The use of magnetic devices in the collection and analysis of the preverbal utterances of an infant—A. W. Lynip

VOLUME 45—January-June, 1952

1. Japanese-American personality and acculturation—W. Caudill
2. A statistical study of the Freudian theory of levels of psychosexual development—C. A. Barnes
   Personality characteristics of selected disability groups—D. N. Wiener

Genetic Psychology Monographs *(continued)*

VOLUME 46—July-December, 1952

1. The relationship of social status, intelligence, and sex of ten- and eleven-year-old children to an awareness of poverty—F. J. Estvan
An empirical study of the castration and Oedipus complexes—S. M. Friedman
2. The relationship between projective test scoring categories and activity preferences—M. M. Schwartz
A comparison of formal and content factors in the diagnostic testing of schizophrenia—M. Sherman

VOLUME 47—January-June, 1953

1. Ability and accomplishment of persons earlier judged mentally deficient—D. C. Charles
Variations in the consistency of the behavioral meaning of personality test scores—M. Kornreich
2. Some child-rearing antecedents of aggression and dependency in young children—R. R. Sears, *et al.*
Symptom correlates for descriptive diagnosis—J. R. Wittenborn, *et al.*

VOLUME 48—July-December, 1953

1. Age and mental abilities: A longitudinal study—W. A. Owens, Jr.
The development of a personality questionnaire for drinkers—P. J. Hampton
Personality and physical disease: A test of the Dunbar hypothesis applied to diabetes mellitus and rheumatic fever—D. H. Crowell
2. Socio-economic contrasts in children's peer culture prestige values—B. Pope
A critical review of the stability of social acceptability scores obtained with the partial-rank-order and the paired-comparison scales—S. A. Witryol and G. G. Thompson
A study of the effects of color on Rorschach responses—G. G. Brody

VOLUME 49—January-June, 1954

1. Factors underlying major reading disabilities at the college level—J. A. Holmes
Parent behavior toward first and second children—J. K. Lasko
2. Social-status and intelligence: An experimental study of certain cultural determinants of measured intelligence—E. A. Haggard
Certain determinants and correlates of authoritarianism—S. Siegel
Personalities in faces: I. An experiment in social perceiving—P. F. Secord, W. F. Dukes, and W. Bevan

VOLUME 50—July-December, 1954

1. A study of the relationship between play patterns and anxiety in young children—E. W. Amen and N. Renison
Operational exploration of the conceptual self system and of the interaction between frames of reference—M. Edelson and A. E. Jones
Problem solving: A statistical description of some relationships between organismic factors and selected response measures—N. A. Fattu, E. Kapos, and E. V. Mech
2. The relation of cortical potentials to perceptual functions—C. Chyatte
The import for clinical psychology of the use of tests derived from theories about infantile sexuality and adult character—D. W. Miles
Measuring personality in developmental terms: The Personal Preference Scale.—M. H. Krout and J. K. Tabin

VOLUME 51—January-June, 1955

1. Some relations between techniques of feeding and training during infancy and certain behavior in childhood—A. Bernstein
The expression of personality in drawings and paintings—L. H. Stewart
2. Negative stereotypes concerning Americans among American-born children receiving various types of minority-group education—J. A. Fishman
The Lincoln-Oseretsky Motor Development Scale—W. Sloan

VOLUME 52—July-December, 1955

1. Some personality correlates of sex, sibling position, and sex of sibling among five- and six-year-old children—H. L. Koch
A quantitative Rorschach assessment of maladjustment and rigidity in acculturating Japanese Americans—G. DeVos
Measurement of authoritarianism and its relation to teachers' classroom behavior—H. M. McGee
2. The formal aspects of schizophrenic verbal communication—B. Mirin
A study in an aspect of concept formation, with subnormal, average, and superior adolescents—H. N. Hoffman
Traumatic avoidance learning: Acquisition and extinction in dogs deprived of normal peripheral autonomic function—L. C. Wynne and R. L. Solomon

VOLUME 53—January-June, 1956

1. As the psychiatric aide sees his work and problems—F. L. Wells, M. Greenblatt, and R. W. Hyde
An investigation of avoidance, anxiety, and escape behavior in human subjects as measured by action potentials in muscle—J. D. Brothers
Spread of effect: A critical review—M. H. Marx
2. Stress, fantasy, and schizophrenia: A study of the adaptive processes—O. J. B. Kerner
The attitude structure of the individual: A Q-study of the educational attitudes of professors and laymen—F. N. Kerlinger

VOLUME 54—July-December 1956

1. A study of personality differences between middle and lower class adolescents: The Szondi Test in culture-personality research—L. Rainwater
The assessment of parental identification—S. W. Gray and R. Klaus
2. The influence of social context on impulse and control tendencies in preadolescents—G. H. Zuk
Tender-mindedness versus tough-mindedness in psychology: A reexamination—H. Winthrop
A method for the comparison of groups: A study in thematic apperception—L. C. Schaw and W. E. Henry

VOLUME 55—January-June, 1957

1. Academic performance and personality adjustments of highly intelligent college students—B. M. Horrall
The use of the Vineland Social Maturity Scale in the planning of an educational program for non-institutionalized low-grade mentally deficient children—M. G. Gottsegen
2. The structure and origin of the anal character—H. Beloff
Free expression of adolescents' interests—M. Amatora
The rôle of mass media and the effect of aggressive film content upon children's aggressive responses and identification choices—R. S. Albert
Interest in persons as an aspect of sex difference in the early years—E. W. Goodenough

VOLUME 56—July-December, 1957

1. Some stable response determinants of perception, thinking, and learning: A study based on the analysis of a single test—M. L. Simmel and S. Counts
Potentials of age: An exploratory field study—S. L. Pressey
2. Tactile communication—L. K. Frank
The use of a filmed puppet show as a group projective technique for children—M. R. Haworth
The social competence of middle-aged people—R. J. Havighurst

VOLUME 57—January-June, 1958

1. Psychological and cultural problems in mental subnormality: A review of research—S. B. Sarason and T. Gladwin
2. Developmental aspects of discrimination in relation to adjustment—P. London
Muscular tension as a response to psychological stress in rheumatoid arthritis and peptic ulcer—J. A. Southworth
Frustration: Theory and experiment—R. Lawson and M. H. Marx

VOLUME 58—July-December, 1958

1. Emotional aspects of political behavior: The woman voter—E. M. Bennett and H. M. Goodwin
The accuracy of self estimate: A measure of self-concept reality—R. M. Brandt
Personality factors in social mobility: A study of occupationally mobile businessmen—J. C. Abegglen
2. Interaction between authoritarian and nonauthoritarian principals and teachers—P. Lambert
Some factors related to the choice-status of ninety eighth-grade children in a school society—D. Elkins
Some relationships between fantasy and self-report behavior—R. C. Calogeras

Genetic Psychology Monographs *(continued)*

VOLUME 59—January-June, 1959

1. Freshman Rorschachs and college performance—G. R. Schmeidler, M. J. Nelson, and M. Bristol
   Perceptual aspects of attributed movement—R. W. Coan
   Men and women: Personality patterns and contrasts—E. M. Bennett and L. R. Cohen
2. Personality factors in mothers of cerebral palsied children—G. Boles
   Separatism and integrationism: A social-psychological analysis of editorial content in New York newspapers of three American minority groups—J. A. Fishman and G. S. Fishman
   Self, role, and satisfaction—A. L. Brophy

VOLUME 60—July-December, 1959

1. The constancy of personality ratings over two decades—R. D. Tuddenham
   Mother-son identification and vocational interest—L. H. Stewart
   A study of mother-child relationships in the emotional disorders of children—M. J. Rosenthal, M. Finkelstein, E. Ni, and R. E. Robertson
   On the trail of the wolf-children—W. F. Ogburn and N. K. Bose
2. Measuring the mental health of normal adults—R. F. Peck
   Parental identification in young children—W. Emmerich
   The problems of Jewish religious youth—B. M. Levinson

VOLUME 61—January-June, 1960

1. Identifying socially maladjusted school children—A. A. Adams
   Social interaction in an interracial nursery school—H. W. Stevenson and N. G. Stevenson, *et al.*
   Interests of pre-adolescent boys and girls—M. Amatora
   Educational and sex differences in the organization of abilities in technical and academic students in Colombia, South America—J. F. Filella
2. Some themes in the personalities of German men—L. Rainwater
   Right-left gradients in body image, body reactivity, and perception—S. Fisher
   Longitudinal survey of child Rorschach responses: Younger subjects two to 10 years—L. B. Ames
   Personality patterns related to occupational roles—M. Siegelman and R. F. Peck

VOLUME 62—July-December, 1960

1. The reading process: A descriptive, interdisciplinary approach—M. P. Langman
   Reactions to stress as a function of the level of intelligence—S. M. H. Zaidi
   Rorschach diagnosis by a systematic combining of content, thought process, and determinant scales—P. A. Bower, R. Testin, and A. Roberts
   Longitudinal survey of child Rorschach responses: Older subjects aged 10 to 16 years—L. B. Ames
2. Twenty years of shock therapy in America, 1937-1956: An annotated bibliography—N. H. Pronko, R. Sitterly, and K. Berg
   The quantitative analysis of parent behavior toward psychotic children and their siblings—E. M. Donnelly

VOLUME 63—January-June, 1961

1. Make a sentence test: An approach to objective scoring of sentence completions—E. F. Borgatta
   Some aspects of the relationship between perception and motility in children—E. P. Rothman
   Some psychophysical aspects of mental illness in children—P. H. Berkowitz
   Conceptual thinking in schizophrenic children—G. Friedman
2. Some maternal influences on children's personality and character—J. C. Finney
   A study of concept formation in normals, mental defectives, and brain-damaged adults—R. W. Zaslow
   An investigation of trait relationships among six-year-old children—S. A. Avakian

VOLUME 64—July-December, 1961

1. Coping patterns of preschool children in response to intelligence test demands—A. Moriarty
   Studies of individual and paired interactional problem-solving behavior of rats: II. Solitary and social controls—A. Baron and R. A. Littman
2. Studies of role performance—J. H. Mann
   A psychological assessment of professional actors and related professions—R. Taft
   A cross-cultural study of menstrual taboos—W. N. Stephens
   The biological roots of creativity—H. Gutman

VOLUME 65—January-June, 1962

1. A Rorschach study of the development of personality structure in white and Negro children in a Southeastern community—A. C. Price
   Measuring personality patterns of women—J. Loevinger
   Home interests in early adolescence—M. Amatora
   Social and emotional adjustment during adolescence as related to the development of psychosomatic illness in adulthood—L. H. Stewart
2. A systematic study of interaction process scores, peer and self-assessments, personality and other variables—E. F. Borgatta
   Children's conceptions of the content and functions of the human body—E. Gellert

VOLUME 66—July-December, 1962

1. Ocular-manual laterality and perceptual rotation of literal symbols—I. Flescher
   Parental child-rearing attitudes and their correlates in adolescent hostility—S. B. Chorost
   Some explorations into the nature of anxieties relating to illness—P. R. Robbins
   Developmental-interactive aspects of hypnosis: Some illustrative cases—J. R. Hilgard and E. R. Hilgard
2. Teaching a two-year-old to read: An experiment in early childhood learning—W. Fowler
   Achievement gains or losses during the academic year and over the summer vacation period: A study of trends in achievement by sex and grade level among students of average intelligence—K. M. Parsley, Jr., and M. Powell

VOLUME 67—January-June, 1963

1. Verbatim dialogue of a mother and child in therapy—C. E. Moustakas
   A study of self-concept: Delinquent boys' accuracy in selected self-evaluations—W. E. Amos
   An integrating study of the factor structure of adult attitude-interests—R. B. Cattell and J. Horn
   Social factors in mental disorders: A contribution to the etiology of mental illness—W. G. Cobliner
2. Conceptual preferences associated with choice of quantitative and nonquantitative college majors—J. Gilbert
   A study of the concept of the feminine role of 51 middle-class American families—A. Steinmann

VOLUME 68—July-December, 1963

1. Brain dynamism as reflected in illusions—G. K. Yacorzynski
   The nature and measurement of components of motivation—R. B. Cattell, J. A. Radcliffe, and A. B. Sweney
2. Sharing in preschool children as a function of amount and type of reinforcement—W. F. Fischer
   The Gesell Incomplete Man Test as a measure of developmental status—L. B. Ames and F. L. Ilg
   Psychological and sociological factors in prediction of career patterns of women—M. C. Mulvey
   Functionalism and its historical significance—R. Harrison

PUBLICATION IN THE JOURNAL OF PSYCHOLOGY

The Journal of Psychology provides a unique service to authors. Articles accepted for publication in this periodical appear in print with the least possible amount of delay. In line with this policy, all steps of the acceptance procedure are maximally streamlined, and the author's work is presented with a bare minimum of editing or restriction of any kind. *The Journal of Psychology* is the ideal medium for presentation of timely professional material.

Authors are requested to exercise unusual care in the preparation of papers for this journal. Accuracy, grammar, and style should be checked and rechecked before a clean version of the manuscript is submitted. Careful attention should be paid to the General Instructions on the back cover of each issue and to the announcement on the inside front cover.

Authors are particularly requested to note the fact that all parts of the manuscript—including references, footnotes, and tables—must be double-spaced.

$30.00 per annum
$15.00 per volume
$7.50 per half volume

BIMONTHLY
Three volumes per year
Immediate publication

January, 1968
Volume 68, First Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a bilateral contract, and may not be resold for a period of five years from the date of publication.

JANUARY, 1968

(Manuscripts are printed in the order of final acceptance)

Fakability of the Sixteen Personality Factor Questionnaire, Form C . . . 3
By John R. Braun and Dolores LaFaro

The case for efferent change during prism adaptation 9
By Gene Lester

The influence of sex and race on process-reactive ratings of schizophrenics . 15
By Ellen A. Lane

The development of some procedures for the assessment of risk 21
By Henry S. G. Cutter and Frederick H. Heilizer

Rationality, motivation, and extinction 33
By Roy Gladstone and Mercedes Miller

Induction of action and attitude change: The function of role-self conflicts and levels of endorsement 39
By Phyllis G. Oram

(OVER)

Copyright © 1968 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved

Published at 35 New St., Worcester, Mass.
Second-class postage paid at Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts

Reading ease as a factor in improved communication effectiveness . . . 49
By Charles V. Keeran, Jr. and Graham B. Bell

Brightness and saturation changes resultant from temporal manipulation of spectral photic input for normal and color-deficient observers . . . 55
By Richard J. Ball and S. Howard Bartley

Cohesion in Marine recruit platoons 63
By LCDR Paul D. Nelson, MSC, USN, and CDR Newell H. Berry, MSC, USN

The comparability of WAIS and WISC subtest scores and *IQ* estimates . . 73
By Mohammed Y. Quereshi

Predicting the full scale WAIS *IQ* of Army basic trainees 83
By Joseph DeLuca

Affective correlates of reading comprehension 87
By Wendell W. Weaver, William F. White, and Albert J. Kingston, Jr.

An analysis of the use of analogies and metaphors in science 97
By Roy Dreistadt

Personality correlates of withdrawal from smoking 117
By Arthur Resnikoff, Paul G. Schauble, and Robert H. Woody

Relevant and irrelevant cues in spatial reversal learning 121
By Robert G. Weyant

A mathematical representation of time-variant quantitative behavior . . 129
By John Voevodsky

Authoritarianism, locus of control, and risk taking 141
By Robert A. Baron

Psychosexual conflict in adolescent enuretics 145
By Joseph DeLuca

Symptom patterns in psychiatric patients: A goal-directed approach to psychiatric symptomatology 151
By Alvin Mahrer, Patricia Stewart, John Horn, and Dennis Lind

Form equivalence of the revised Visual Retention Test in mental retardates: An extension 159
By James A. Rice

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 3-7.

FAKABILITY OF THE SIXTEEN PERSONALITY FACTOR QUESTIONNAIRE, FORM C*

Department of Psychology, University of Bridgeport

JOHN R. BRAUN AND DOLORES LAFARO[1]

A. INTRODUCTION

The Sixteen Personality Factor Questionnaire (16PF) is a factor-analytically derived inventory yielding scores on 16 basic dimensions (2). Form C is a short form which ". . . tends most frequently to be used in less satisfactory testing situations, e.g., mass testing of employees, where the most gross attempts at faking, under strong motivation to obtain jobs, etc., are most likely to occur" (2, p. 6). To help cope with faking, a Motivational Distortion (MD) score may be derived. The MD scale consists of those items that have shown the greatest change in response between an anonymous, presumably frank condition of test administration and a condition in which *S*s were motivated to create favorable impressions of themselves (2). Certain corrections of scores on two personality factors may be made if the MD score exceeds 12 points. However, the manual (2) implies that faking may not be much of a problem with this inventory. It is suggested that the opposite poles of the dimensions on factors A, E, F, I, M, N, and Q 1 seem equally desirable to most *S*s. Similarly, with factors C, G, L, O, Q 3, and Q 4 where social desirability considerations might be expected to operate, ". . . no systematic motivational distortion has yet been experimentally found" (2, p. 6).

The 1966 *Standards for Educational and Psychological Tests and Manuals* (4, p. 24) lists the following as "very desirable": "For any personality measure, evidence should be presented on the extent to which scores are susceptible to an attempt by the examinee to present a false or unduly favorable picture of himself." The 16PF manual falls short of meeting this recommendation. Although the problem of response distortion is discussed, no evidence in the form of specific faking studies is provided. The present study therefore investigated the effects of a faking set on the 16 basic dimensions of Form C, and also, the effectiveness of the MD scale in detecting faking.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on September 29, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The authors are indebted to Arthur Salman for his kind assistance in obtaining the data reported in this study.

B. METHOD

*S*s were students in introductory psychology classes. Form C was administered in regular class sessions with either regular instructions (Control Group, $N = 67$), or instructions to fake (Experimental Group, $N = 87$). Classes were assigned to the control or experimental condition in such a way as to equalize the factors of class meeting time and day between the conditions.

Control *S*s received the following instructions:

> Today we shall be taking a personality questionnaire. General instructions for taking the test are presented on the cover of the question booklet. No marks should be made in the question booklet; instead, all answers go on the special answer sheet. You do not need to put your name on the answer sheet—just your sex, M or F. Although we cannot tell you anything about this study at this time, we will be able to give you some information next week, and it is anticipated that further information will be made available later in the semester. Since we shall be gathering data throughout the remainder of this week in other sections, please do not discuss the study with anyone this week. Your cooperation is essential. Now please read the instructions on the question booklet cover.

Experimental *S*s received the above instructions, and also the following:

> One of the problems with personality questionnaires is that people may try to slant their answers so as to give a good impression of themselves. That is, instead of giving answers that reflect the way they actually feel or act, they give answers that they believe will place them in a good light. They may try to fake their answers somewhat.
>
> Personality inventories vary in the extent to which such deliberate faking is possible. With some it is very easy; with others quite difficult. We are studying how the 16PF is in this regard. Contrary to the instructions given on the cover of the booklet, we want you to answer not necessarily with your true feelings, but rather in the way that you would if you were actually going to try to fake—to beat the test.
>
> However, we must warn you that the test has certain features by which it attempts to detect faking. Answer the test so that you make the best impression you can, but be careful to do it in such a way that your faking cannot be detected.

Of the 67 control *S*s, 30 were male and 37 were female; while for the 87 experimental *S*s, 48 were male and 39 were female.

C. RESULTS AND DISCUSSION

Means, standard deviations, and results of significance tests (*t* tests) are presented in Table 1. Differences between control and experimental means were significant for eight of the 16 test factors. The direction of these significant differences was such that *S*s, when faking, appeared mature rather than emotional, enthusiastic rather than glum and silent, conscientious rather than casual, adventurous rather than timid, trustful rather than suspecting, simple rather than sophisticated, self-controlled rather than uncontrolled, and stable rather than tense.

The test manual suggested that the opposite poles of factors A, E, F, I,

TABLE 1

Means and Standard Deviations for Regular and Faked Administrations

	Control ($N = 67$)		Experimental ($N = 87$)	
Variable	*M*	*SD*	*M*	*SD*
Factor A—aloof *versus* warm, outgoing	7.61	2.30	7.49	2.58
Factor B—dull *versus* bright	4.76	1.25	4.70	1.21
Factor C—emotional *versus* mature	6.78	2.22	7.64*	2.47
Factor E—submissive *versus* dominant	5.69	2.72	5.02	2.64
Factor F—glum, silent *versus* enthusiastic	7.00	2.75	8.06*	2.66
Factor G—casual *versus* conscientious	5.09	2.52	6.37**	2.59
Factor H—timid *versus* adventurous	5.90	2.45	6.76*	2.30
Factor I—tough *versus* sensitive	6.31	2.88	6.03	2.73
Factor L—trustful *versus* suspecting	6.60	2.09	5.74*	2.42
Factor M—conventional *versus* eccentric	6.72	2.02	6.15	1.88
Factor N—simple *versus* sophisticated	5.61	1.81	4.93*	2.15
Factor O—confident *versus* insecure	4.66	1.96	4.44	1.79
Factor Q 1—conservative *versus* experimenting	6.24	2.10	5.78	2.29
Factor Q 2—dependent *versus* self-sufficient	6.99	1.89	7.18	1.77
Factor Q 3—uncontrolled *versus* self-controlled	5.73	2.32	7.25**	2.43
Factor Q 4—stable *versus* tense	6.31	2.39	5.38*	2.86
Motivational distortion (MD)	6.93	2.82	8.17**	2.89

* Significantly different from Control mean at either .05 or .02 level.

** Significantly different from Control mean at either .01 or .001 level.

Note: For the factor scores, low scores are associated with the term listed first (e.g., Factor A—aloof); high scores with the second term (e.g., Factor A—warm, outgoing).

M, N, and Q 1 seem equally desirable to most people, and that, therefore, faking was not a problem with these factors. Our results lend partial support to this view, since the faking set did not result in significantly altered scores on factors A, E, I, M, and Q 1. However, contrary to the manual, significant effects of the faking set were found with factors F and N.

The manual indicated also that no evidence had been found for motivational distortion effects with factors C, G, L, O, Q 3, and Q 4. Contrary to this, our results show significant effects of the faking set for all of these dimensions except factor O.

The experimental group mean Motivational Distortion score was significantly higher than that for the control group ($p < .01$). However, only five per cent of the experimental group MD scores were as great as 13 (the value suggested as crucial by the manual), while three per cent of the control group scores reached this value. The difference between the proportion of experimental and control group *S*s with MD scores $\geqq 13$ was not significant. Similar findings were obtained with MD cutoff values of 12, 11, and 10. It is only when we drop down to an MD cutoff point as low as 9 that the difference between control and experimental group is significant ($p < .05$). The MD scale was thus of little value in detecting faking under the conditions of the present study.

How representative are the present conditions of those under which persons ordinarily take Form C? Our instructions to experimental *S*s warned that there were ways of detecting faking. Similar information routinely appears on the cover of the Form C booklet: "Never give an untrue answer about yourself because you think it is the 'right thing to say.' There are ways of detecting such unfair answers." According to Lovell (3), "The public seems well informed of the basic principles underlying the rather crude devices presently in existence for the detection of faking." In the sense of *S*s' knowing that techniques exist for detecting faking, there is little reason to believe that our situation was unrepresentative of that involved in typical use of Form C. Whether our *S*s vary in ability to fake successfully from the general run of persons who take Form C in personnel settings can be only a matter of conjecture at present.

It was indicated earlier that corrections of scores on two personality factors may be made if the MD score exceeds 12 points. This involves adding one raw score point to Q 2 and subtracting one raw score point from H (1, p. 9). Our data confirm the value of this suggestion for H, since experimental *S*s earned significantly higher scores than did control *S*s. Subtracting a point pushes the faked scores back in the direction of control scores. However, our data do not support the suggestion that one point be added to Q 2. Our experimental *S*s were higher (although not significantly so) than control *S*s on Q 2. Adding a point to Q 2 pushes scores further in the faked direction.

In conclusion, despite test manual suggestions to the contrary, scores on eight of the 16PF dimensions were significantly affected by the faking set. The Motivational Distortion scale was of quite limited usefulness in discriminating between control and faked protocols.

D. SUMMARY

The effects of faking instructions on Form C of the Sixteen Personality Factor Questionnaire were investigated. Differences between control ($N = 67$) and experimental ($N = 87$) means were significant for eight of the 16 basic test dimensions. Although the mean Motivational Distortion score was significantly higher for experimental than for control *S*s, the MD scale was of little use in detecting faking.

REFERENCES

1. CATTELL, R. B. Handbook Supplement for Form C of the Sixteen Personality Factor Questionnaire (2nd ed.). Champaign, Ill.: Inst. for Personality and Ability Testing, 1962.
2. CATTELL, R. B., & EBER, H. W. Handbook for the Sixteen Personality Factor Questionnaire. Champaign, Ill.: Inst. for Personality and Ability Testing, 1957.
3. LOVELL, V. R. The human use of personality tests: A dissenting view. *Amer. Psychologist,* 1967, **22**, 383-393.
4. STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTS AND MANUALS. Washington: American Psychological Association, 1966.

Department of Psychology
University of Bridgeport
Bridgeport, Connecticut 06602

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 9-13.

THE CASE FOR EFFERENT CHANGE DURING PRISM ADAPTATION*

Department of Psychology, Wheaton College

GENE LESTER

A. INTRODUCTION

Harris (16) has outlined a theory of adaptation to distorted vision. His approach, which involves a change in sensed position of body parts, stresses the inflexibility of the visual system as shown by Hubel and Wiesel (20). Harris suggests that felt positions of body parts are developed and modified according to information from the visual system. Although Harris does not suggest mechanisms of body position identification, recognizing the difficulties in being sure of such mechanisms, he refers to changes in apparent body position as proprioceptive changes; the term "proprioceptive" could apply to either afferent or efferent processes. Kern, Lotto, and Morant (23) have proposed a similar explanation for prism adaptation and have demonstrated the interdependence of apparent changes in visual vertical and postural vertical of the head.

Since an understanding of these adaptation phenomena and the development of associated theory depend on the contributions of the afferent and efferent systems, there is a need to evaluate what may be said about possible relationships. The present paper reviews the evidence concerning the dependence of changes in apparent body position on changes in efference and afference.

B. AFFERENT CHANGES

Mechanisms for change in peripheral proprioceptive afferent activity are known to exist. Leksell (26) found in ventral root fibers going to skeletal muscle a number of small fibers (the gamma efferent system) for which no motor function could be identified. He concluded that they regulate sensory activity by their influence on the muscle spindles. Kuffler *et al.* (24) and Hunt and Kuffler (21) confirmed this and stated that "this efferent system appears well suited to act as a peripheral regulator of the proprioceptive muscle spindle mechanism" (24, p. 54). Eldred *et al.* (5) found that deafferenta-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 2, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

tion did not diminish spindle activity, but de-efferentation produced a sudden drop in and regularization of spindle discharge frequency. Granit and Kaad (13) later showed that the muscle spindles were facilitated by stimulation of the brain stem and other areas; inhibition could also be produced by stimulation of higher nervous centers. Hongo (19) found that the effects of such stimulation on the annulo-spiral and flower-spray endings were the same.

C. THE IMPORTANCE OF EFFERENCE

There is considerable evidence that afference is not necessary as an indicator of body part position. Taub, Ellman, and Berman (29) conditioned a grasp response in a deafferented primate limb and concluded that limb movement was less under the control of afferent information than had been thought. Gorska and Jankowska (8, 9, 10, 11, 12, 22) have worked on motor performance after deafferentation of a limb in rats, cats, and dogs. These workers found that animals were able to perform previously conditioned instrumental movements with a deafferented limb. The movements were somewhat simplified and reduced, but did not need to be retrained after operation. No gradual deterioration occurred after deafferentation. New instrumental responses could be conditioned without allowing the animal to see the limb. Gorska and Jankowska felt that their results were probably due to the contribution of information about a movement by initiation of the movement at a motor center. However, information coming from other parts of the body could act as a substitute feedback mechanism in the absence of normal afference.

The experiments discussed above were all performed with subhuman subjects. Laszlo (25), using human *S*s, found a significant loss in key-tapping ability during loss of kinaesthesis from a pressure nerve block, although muscle power was not seriously affected. Thus, conclusions about the redundancy of afference may not necessarily be applicable to humans. However, afference might be necessary only for relatively fine tasks like tapping with one finger, and not for coarser activities like moving an entire limb or grasping.

For purposes of analysis of change in spatial localization, it is especially important to know what role afference and efference play in information about eye position. Brindley and Merton (2) have shown that there is probably no afferent basis for registration of eye position. Merton (27) found that the accuracy of directing the hand and the eyes in the dark was very poor. He concluded that proprioceptors contribute little to reaching or visual fixation. Festinger and Canon (6) showed that fixation of a light which appeared and disappeared suddenly gave a more accurate idea of the light's position than did tracking the light to a point where it disappeared. This was interpreted as meaning that the best piece of information about an eye's position is the registration of a command for it to move to that position. Later experiments by Festinger and his co-workers gave results that were "consistent with the theoretical position that the efference and efferent readiness activated by visual input help determine the visual perception of contour" (7, p. 1).

It was suggested by Breinin (1) that the electrical activity of detached eye muscles makes it necessary to posit some afferent activity, though possibly not of a conscious sensory type. Ludvigh [quoted in Davson (4)] felt that the many afferent fibers in the extraocular muscles serve only as indicators of metabolic state and of very gross eye position. Christman and Kupfer (3) suggested that the role of the extraocular muscle spindles is to increase the sensitivity of the muscles to stretch during fixation. This would improve reflex control of fine movements during fixation, but would yield no information about eye position.

D. Discussion

These findings suggest that changes in the position sense may not be based on changes in afference from proprioceptors. Another possible mechanism lies in changed monitoring of efference. Motor outflow going to the two sides of the body could be monitored and compared. Complex interactions would involve monitoring of outflow, comparison, transfer of correction signals back to the motor areas, and monitoring of the changed signals. Changes in the sensed position of body parts might then depend on the ratio of outflow to the two sides of the body which the monitor accepted as 1. If a ratio of 1.1, with the greater outflow going to the right side of the body, were accepted by the monitor as a ratio of 1, the body would be felt as symmetrically oriented when some torsion to the right was objectively present. Objective symmetrical orientation would be felt as torsion to the left.

The responsibility of monitoring of efferent signals for knowledge about movements was originally suggested by Helmholtz (18) and has been supported by Gregory (15). The mechanism of a change in monitoring of the efferent signals is not an easy thing to determine, since the monitor itself is hypothetical. Hein and Held (17) have suggested a process for correlating efference and afference which includes a correlation storage center and a comparator, but have not really offered explanations of the processes of monitoring and comparison.

Changes in efference may be most easily studied in the changed motor patterns of *S*s who are experiencing visual distortions. A paradigm of these motor changes occurs in *S*s who have received vestibular stimulation, which leads to apparent displacement of a visual target (14). When instructed to stand facing straight ahead and with arms extended straight in front, the blindfolded *S*s exhibit a posture known as the "discus-thrower's position." Deceleration from rotation to the right produces the following movements: turning of head, trunk, and arms to the right; raising of the left arm and lowering of the right arm; and extension of the left leg, with flexion of the right leg [Wendt, cited by Stevens (28)].

Similar changes in motor patterns appear to take place in *S*s who are wearing distorting prisms. *S*s who are undergoing adaptation to tilting prisms display obvious postural changes (personal observations). The entire body is tilted in the direction of prism tilt while *S* is walking. The foot on the side in the direction of tilt strikes the ground forcefully, while the other foot is extended awkwardly to one side and skims the ground lightly at each step.

Although investigations of these changed motor patterns cannot give us information about the process of monitoring motor outflow, they may well help us to infer what sorts of efferent changes are taking place during adaptation to visual distortions.

E. SUMMARY

Proprioceptive-change theories of adaptation to distorted vision do not make it clear whether the postulated changes occur in the afferent or the efferent system. Evidence from the literature leads to the suggestion that the changes are in the efferent system.

REFERENCES

1. BREININ, G. M. Electromyographic evidence for ocular muscle proprioception in man. *Arch. Ophthal.,* 1957, **57**, 176-180.
2. BRINDLEY, G. S., & MERTON, P. A. The absence of position sense in the human eye. *J. Physiol.,* 1960, **153**, 127-130.
3. CHRISTMAN, E. H., & KUPFER, C. Proprioception in extraocular muscles. *Arch. Ophthal.,* 1963, **69**, 824-829.
4. DAVSON, H., *Ed.* The Eye: Vol. III. Muscular Mechanisms. New York: Academic Press, 1962.
5. ELDRED, E., GRANIT, R., & MERTON, P. A. Supraspinal control of the muscle spindles and its significance. *J. Physiol.,* 1953, **122**, 498-523.
6. FESTINGER, L., & CANON, L. K. Information about spatial location based on knowledge about efference. *Psychol. Rev.,* 1965, **72**, 373-384.
7. FESTINGER, L., ONO, H., BURNHAM, C. A., & BAMBER, D. Efference and the conscious experience of perception. *J. Exper. Psychol. Monog.,* 1967, **74**, No. 4 (Whole No. 637).
8. GORSKA, T., & JANKOWSKA, E. Instrumental conditioned reflexes of the deafferented limb in cats and rats. *Bull. Acad. Pol. Sci.,* 1959, **7**, 161-164.
9. ———. The effect of deafferentation on the instrumental conditioned reflexes established in dogs by reinforcing passive movements. *Bull. Acad. Pol. Sci.,* 1960, **8**, 527-530.

10. ———. The effect of deafferentation on instrumental (Type II) conditioned reflexes in dogs. *Acta Biol. Exper.*, 1961, **21**, 219-234.

11. ———. The effect of deafferentation of a limb on instrumental reflexes. In E. Gutmann (Ed.), *Proceedings of the Conference on Central and Peripheral Mechanisms of Motor Functions, Czechoslovak Academy of Science.* Prague: P. Hnik, 1963.

12. Gorska, T., Jankowska, E., & Kozak, W. The effect of deafferentation on instrumental (Type II) cleaning reflex in cats. *Acta Biol. Exper.*, 1961, **21**, 207-217.

13. Granit, R., & Kaada, B. R. Influence of stimulation of central nervous structures on muscle spindles in cat. *Acta Physiol. Scand.*, 1953, **27**, 130-160.

14. Graybiel, A., Clark, B., MacCorquodale, K., & Hupp, D. Role of vestibular nystagmus in the visual perception of a moving target in the dark. *Amer. J. Psychol.*, 1946, **59**, 259-266.

15. Gregory, R. L. Eye movements and the stability of the visual world. *Nature*, 1958, **122**, 1214-1216.

16. Harris, C. S. Perceptual adaptation to inverted, reversed, and displaced vision. *Psychol. Rev.*, 1965, **72**, 419-444.

17. Hein, A., & Held, R. A neural model for labile sensorimotor co-ordinations. In *Biological Prototypes and Synthetic Systems (Vol. I)*. New York: Plenum Press, 1962.

18. Helmholtz, H. von. Helmholtz's Treatise on Physiological Optics. J. P. C. Southall, *Ed.* Washington, D.C.: The Optical Society of America, 1925.

19. Hongo, T., & Shimazu, H. Centrifugal modifications of discharge rates of primary and secondary endings of muscle spindles in the hind limb of the cat. *J. Neurophysiol.*, 1965, **28**, 724-741.

20. Hubel, D. H., & Wiesel, T. N. Receptive fields in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiol.*, 1963, **26**, 994-1002.

21. Hunt, C. C., & Kuffler, S. W. Further study of efferent small-nerve fibers to mammalian muscle spindles. Multiple spindle innervations and activity during contraction. *J. Physiol.*, 1951, **113**, 283-297.

22. Jankowska, E. Instrumental scratch reflex of the deafferented limb in cats and rats. *Acta Biol. Exper.*, 1959, **19**, 233-248.

23. Kern, A., Lotto, D., & Morant, R. B. Prism-induced tilt aftereffects and changed felt head position. Paper given at EPA, Boston, 1967.

24. Kuffler, S. W., Hunt, C. C., & Quilliam, J. P. Function of medullated small-nerve fibers in mammalian ventral roots: efferent muscle spindle innervation. *J. Neurophysiol.*, 1951, **14**, 29-54.

25. Laszlo, J. I. The performance of a simple motor task with kinaesthetic sense loss. *Quart. J. Exper. Psychol.*, 1966, **18**, 1-8.

26. Leksell, L. The action potential and the excitatory effects of the small ventral root fibers to skeletal muscle. *Acta Physiol. Scand.*, 1945, **10**, Supp. 31.

27. Merton, P. A. The accuracy of directing the eyes and the hand in the dark. *J. Physiol.*, 1961, **156**, 555-577.

28. Stevens, S. S. Handbook of Experimental Psychology. New York: Wiley, 1951.

29. Taub, E., Ellman, S. J., & Berman, A. J. Deafferentation in monkeys: Effect on conditioned grasp response. *Science*, 1966, **151**, 593-594.

Department of Psychology
Wheaton College
Norton, Massachusetts 02766

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 15-20.

THE INFLUENCE OF SEX AND RACE ON PROCESS-REACTIVE RATINGS OF SCHIZOPHRENICS* [1]

Psychology Department, Case Western Reserve University

ELLEN A. LANE

A. INTRODUCTION

If intelligence test performance is influenced by the social class of the individuals tested, to the disadvantage of those from low socioeconomic backgrounds and from other cultures (5), it should not be surprising to find that process-reactive ratings of schizophrenics, emphasizing sexual and social behavior, are also not culture-free. Construction of culture-free scales is no more important for estimating intelligence than for estimating the severity, personality growth, and prognosis of schizophrenics.

There is evidence that the process-reactive scales in present use are not free of social class differences. Baxter and Arthur (1), for example, found that interparental conflict is associated with good premorbid adjustment for schizophrenics from the middle class but is associated with poor premorbid adjustment for schizophrenics from the lower class. They conclude that "the behavioral differences thought to be related to schizophrenia could in fact be a function of social class, or perhaps a joint function of both social class and premorbid adjustment" (1, p. 273).

The hypothesis presented in the present study is that, in addition to the possible contamination by differences in social class values, the behavior pertinent to process-reactive ratings of schizophrenia is unduly influenced by differences in sex and race. With the increased use of the process-reactive dimension as a measure of prognosis, it is important to isolate those factors extraneous to schizophrenic development that affect the validity of a measurement which can determine who will receive the best available treatment and who will not.

B. METHOD

Subjects were 200 adult schizophrenics: 50 Negro males, 50 white males, 50 Negro females, and 50 white females. Subjects were chosen from consecu-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 9, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The research reported in this paper is supported by Grant M-5186 from the National Institute of Mental Health, United States Public Health Service, Department of Health, Education, and Welfare.

tive first admission files of several public hospitals in Cleveland until the desired numbers of each sex and race were obtained. All first-admission schizophrenics were included who were over 16 and under 40 years of age on first admission and who were born in Ohio. A large majority of both white and Negro subjects came from low socioeconomic backgrounds.

The Symptom Checklist (12), which is reported to correlate highly with Phillips Prognostic Rating Scale (15), was used to divide subjects into "process" and "reactive" categories. The ratings were done independently by three trained psychology majors who did not select the subjects to be rated and who were not aware of the purpose of the study.

The scale itself consists of a checklist of behavior present upon first hospitalization as recorded by a psychiatrist in an initial interview. With a simple formula, schizophrenics are rated according to the relative number of behaviors considered aggressive toward one's self (reactive) and behavior considered either withdrawn or aggressive toward others (process). Process subjects were divided from reactive subjects at .40 on a continuum of scores from .00 to 1.00 (reactive).

C. RESULTS

Even though the distinction between "process" and "reactive" was made at a point below the midpoint of the scale, 64 per cent of the subjects of this study turned out to be process schizophrenics, while only 36 per cent were reactive schizophrenics. It is obvious from Table 1 that most of the disproportion was contributed by Negro subjects of whom 75 per cent were considered process and 25 per cent reactive. White schizophrenics were divided more evenly into 53 per cent process and 47 per cent reactive. While the differences in racial composition in the "process" and "reactive" categories are apparent throughout, with a significantly higher percentage of Negroes than whites receiving "process" ratings, these differences were more pronounced between males of the two races than between females.

TABLE 1
NUMBER OF PROCESS AND REACTIVE SCHIZOPHRENICS GROUPED ACCORDING TO RACE AND SEX

Group	*N*	Process	Reactive
Negro	100	75	25
White	100	53	47
Male	100	71	29
Female	100	57	43
Negro male	50	42	8
White male	50	29	21
Negro female	50	33	17
White female	50	24	26

Table 1 also shows that 71 per cent of all males were rated "process," while only 50 per cent of the females received this designation. However, the statistical results in Table 2 show that the male-female difference in ratings, significant for the total group, was not significant for the white population alone. The significantly higher number of "process" ratings for Negro males than for Negro females is not paralleled for white males and white females, in spite of the fact that white females were the only group receiving fewer "process" than "reactive" ratings. Being Negro increases one's chances of being rated "process" but being Negro and male almost assures it. The difference between Negro males and white females on these ratings was well beyond the .001 level of significance on the χ^2 test.

TABLE 2

CHI SQUARE TESTS OF INFLUENCE OF RACE AND SEX ON PROCESS-REACTIVE RATINGS

Group	*N*	χ^2	*p*
Negro *vs.* white	200	10.504	$< .01$
Male	100	6.990	$< .01$
Female	100	3.312	$< .10$
Male *vs.* female	200	4.254	$< .05$
Negro	100	4.320	$< .05$
White	100	1.034	$< .50$

D. DISCUSSION

While the results clearly indicate that race and, to some extent, sex are closely related to process-reactive ratings of schizophrenia, several alternate explanations of these results are possible. If similar patterns of differences between races and sexes are found by other investigators with other measures of process-reactive behavior, the same explanations would apply:

1. *The scale may measure valid differences*. The fact that a disproportionate number of Negroes, especially Negro males, receive "process" ratings does not necessarily mean that the scale is invalid. Instead, a higher proportion of process schizophrenics among Negroes than whites may be a description of reality. In a review of several studies, Chapman and Baxter (3) conclude that process schizophrenics come from mother-dominated homes, while reactive schizophrenics come from father-dominated homes. If lower-class Negro families are mother-dominated (10), and lower-class white families are father-dominated (3), the present finding of more Negro process schizophrenics than white process schizophrenics from the lower socioeconomic levels might be expected. Other authors (9), however, report that mother-dominated homes are typical of lower-class schizophrenics in general.

There is some justification for expecting process-reactive differences between the sexes: one study (15) found that female psychotics tend to turn aggression inward, while male psychotics more often turn aggression outward; another (4) found that male schizophrenics are more withdrawn than female schizophrenics. In addition, more male than female schizophrenics are admitted to hospitals (8). Kantor and Herron (8) present a theory of integrative personality that would account for a higher proportion of female than male reactives.

There still remains a more general question of whether most schizophrenics from the lower class are process schizophrenics, as found in the present study (it will be remembered that a dividing point nearer the "process" end of the scale was chosen, so small was the number of schizophrenics with high "reactive" ratings); or whether reactive schizophrenics come from the lower classes and process schizophrenics from the higher socioeconomic levels, as reported by others (3). One explanation for these different findings is that, rather than representing a valid description of the schizophrenic population, they result from biases in the scales which show up when populations of different racial compositions are used, a possibility discussed next.

2. *The rating scale may measure differences in normal patterns of behavior rather than, or in addition to, differences in schizophrenic behavior.* This hypothesis is the opposite of the first and implies that the rating procedure is largely invalid. The particular rating scale used in this study emphasizes the direction of aggression as a distinguishing mark between process and reactive schizophrenia. Aggression toward others may be the more commonly accepted and encouraged behavior for males than for females, and perhaps also for Negroes than for whites at the lower socioeconomic levels. If so, normal subcultural patterns of behavior are reflected in the results rather than patterns that truly distinguish one kind or degree of schizophrenia from another. There may also be differences in the direction and degree of aggression sanctioned by different socioeconomic levels, but evidence for this pattern among schizophrenics would have to come from studies with subjects from more diverse backgrounds than the present sample.

A way to test whether the differences found between Negro and white and between male and female are more related to culture than to severity of disorder would be to apply the rating scale to personality descriptions of normal control subjects. If the relative positions of nonschizophrenic Negroes and whites, and males and females, are the same as for the schizophrenic group, even if the range of scores for the normal group is higher, subcultural factors related to race and sex could be said to play an important role in the results.

3. *The rating scale may reflect the biases of the interviewing psychiatrist.*

A third possible explanation for the results is one of personal interviewer bias. Hollingshead and Redlich (7) described the presence of hostility and "dislike" on the part of the upper middle-class white psychiatrists toward their lower-class clients. Such feelings during the psychiatric interview could well play a part in the present results. There is also evidence that female schizophrenics are likely to receive more attention and more therapy than male schizophrenics (13), whether because of better prognosis or personal bias is not established.

If psychiatrists perceive Negroes as more threatening and aggressive than whites, and males, especially Negro males, as more threatening than females, the resulting ratings would tend to rank Negroes and males as "process." Further, if psychiatrists' biases, reflected in manner, actions, and questions, bring out hostility in Negroes and in males, the same results would ensue. Different results from the present ones might have occurred had the interviewers been Negro or female, a phenomenon that has been demonstrated with the administration of intelligence tests (2), and one that might be investigated for those interviews, whether by psychiatrists or social workers, that form the basis of many process-reactive rating scales.

Whichever the appropriate explanation, the results strongly suggest that a schizophrenic's race and sex will be related to whether he falls nearer the "process" or "reactive" end of present-day rating scales. Research with other scales than that used in the present study is, of course, needed to substantiate this general conclusion. While the present scale is based on the behavior of persons upon admission, the "presenting symptoms" (and the psychiatrist's perception of them), other scales, such as the Elgin Scale (14) and the Phillips Premorbid History Scale (11), that put greater emphasis on social and sexual history, may be equally, perhaps even more, subject to sexual and racial differences, as they are to socioeconomic factors (3). Even studies that distinguish "process" and "reactive" schizophrenia without the use of scales by equating the two concepts with "chronic" and "acute" are not immune from biased categorization, for length of hospitalization is influenced by social class (7) and may also be influenced by race and sex.

A bias in any method of classification of process and reactive schizophrenics can have important consequences to individuals concerned. Just as many intelligence tests were originally standardized on all white populations, most standard process reactive rating scales were also validated on groups of white psychotics with the apparent result that some of the behaviors and values included may not be equally meaningful criteria of process and reactive schizophrenia for both groups. With the recent surge of research and theory concerned with the "process" and "reactive" concepts (6), it is essential that racial, sexual, cultural, and class factors that are not directly related to the nature of schizophrenia be eliminated as much as possible from the classification process.

E. Summary

Two hundred first-admission schizophrenics were divided into "process" and "reactive" groups with Phillips Symptom Checklist. A significantly higher proportion of Negro than white schizophrenics were rated "process." Male Negroes fared significantly worse than female Negroes. Three alternate ways of understanding the results are presented: their essential validity, bias in the items of rating scales, and personal bias on the part of the interviewers.

References

1. Baxter, J., & Arthur, S. C. Conflict in families of schizophrenics as a function of premorbid adjustments and social class. *Family Process,* 1964, **3**, 273-279.
2. Canady, H. G. The effect of "rapport" on the IQ: A new approach to the problem of racial psychology. *J. Negro Educ.,* 1936, **5**, 209-219.
3. Chapman, L. J., & Baxter, J. C. The process-reactive distinction and patient's sub-culture. *J. Nerv. & Ment. Disor.,* 1963, **136**, 352-359.
4. Check, F. E. A serendipitous finding: Sex roles and schizophrenia. *J. Abn. & Soc. Psychol.,* 1964, **69**, 392-400.
5. Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler, R. Intelligence and Cultural Differences. Chicago, Ill.: Univ. Chicago Press, 1951.
6. Higgins, J., & Peterson, J. C. Concept of process-reactive schizophrenia: A critique. *Psychol. Bull.,* 1966, **66**, 201-206.
7. Hollingshead, A. B., & Redlich, F. C. Social Class and Mental Illness. New York: Wiley, 1958.
8. Kantor, R. E., & Herron, W. G. Reactive and Process Schizophrenia. Palo Alto, Calif.: Sci. & Behav. Books, 1966.
9. Kohn, M. L., & Clausen, J. A. Parental authority behavior and schizophrenia. *Amer. J. Orthopsychiat.,* 1956, **26**, 297-313.
10. The Negro Family, The Case for National Action. Office of Policy Planning and Research, United States Department of Labor, March, 1965.
11. Phillips, L. Case history data and prognosis in schizophrenia. *J. Nerv. & Ment. Dis.,* 1953, **117**, 515-525.
12. Phillips, L., & Rabinovitch, M. S. Social role and patterns of symptomatic behaviors. *J. Abn. & Soc. Psychol.,* 1958, **57**, 30-36.
13. Schofield, W. Psychotherapy: The Purchase of Friendship. Englewood Cliffs, N.J.: Prentice-Hall, 1964.
14. Wittman, P. M. Scale for measuring prognosis in schizophrenic patients. *Elgin State Hospital Papers,* 1941, **4**, 20-33.
15. Zigler, E., & Phillips, L. Psychiatric diagnosis of symptomatology. *J. Abn. & Soc. Psychol.,* 1961, **63**, 69-75.

Psychology Department
Case Western Reserve University
Cleveland, Ohio 44106

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 21-32.

THE DEVELOPMENT OF SOME PROCEDURES FOR THE ASSESSMENT OF RISK*

Veterans Administration Hospital, Brockton, Massachusetts

HENRY S. G. CUTTER AND FREDERICK H. HEILIZER

A. INTRODUCTION

Gambling is a uniquely human activity in which the gambler stakes his fortune on the operation of chance. A gamble frequently involves a dilemma of choice between the high probability of a small payoff or the low probability of a large payoff. Put another way, gamblers have to choose among alternatives which vary along a continuum of risk.

The present paper discusses the development of a series of gambling games that were designed to measure different aspects of risk. The game format and procedures are presented in the sequence in which they were developed partly to maintain historical continuity, but primarily to introduce the reader to the easier games and their potentialities before going on to the more complicated games. The game procedures are modular so that the earlier developments logically fit the later game formulations.

The first gambling game, developed to study risk taking among hospitalized alcoholics (2) evolved from two different techniques of Cohen (1). He asked children to choose between (*a*) a single package of candy, (*b*) a one in five chance of winning five packages of candy, and (*c*) a one in 25 chance of winning 25 packages of candy.

The first choice in one sense is not a gamble at all, since there is no risk of not obtaining the candy. The second and third choices involve increasing risk (*a*) as a function of the decreasing probabilities of winning, and (*b*) as a function of the increasing size of the payoff. The three bets have an equal expected value (EV); they are equivalent in the sense that, if one were to attempt to win each bet for an infinite number of trials, one would win an average of one package of candy per trial. EV, which can be computed by multiplying the probability of winning by the size of the payoff, is held constant in most studies of gambling to eliminate the possibility that one bet be objectively considered more "rational" than another. Cohen's results indicate that his uncomplicated risk measure is capable of assessing interesting differ-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 10, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

ences among children related to age and whether the children are in a vocational or academic school "track."

In other studies Cohen (1) used urns filled with varying numbers of different colored beads to study probability concepts. The latter technique provides a convenient way of concretely representing gambling odds, and a device with which *S*s can actually gamble.

B. GAMES

The initial gambling game used by one of the authors was administered to subjects individually (2). Ten cups, each containing one green bead and from zero to nine red beads respectively, were lined up in that order in front of the subject. The cups were labelled with the number of green and red beads that they contained and the amount of money that they would pay if the subject chose to gamble with that particular cup (Appendix 1). The subjects were allowed to inspect the cups and to make trial draws before choosing a cup. The cup choice number, the index of risk, was recorded by the experimenter. The subject was then allowed to reach blindly into his chosen cup in an attempt to pick out the green bead and win the payoff. With this procedure the experimenter had to exercise care to insure that subjects

APPENDIX 1

GAME NO. I

In Game No. I you hope to win from the instructor. Circle your preferred betting cup number. Make only one choice.

CUP NO.	CONTAINS	PAYS
1	1 green bead 0 red beads	10¢
2	1 green bead 1 red bead	20¢
3	1 green bead 2 red beads	30¢
4	1 green bead 3 red beads	40¢
5	1 green bead 4 red beads	50¢
6	1 green bead 5 red beads	60¢
7	1 green bead 6 red beads	70¢
8	1 green bead 7 red beads	80¢
9	1 green bead 8 red beads	90¢
10	1 green bead 9 red beads	$1.00

could not see the beads when they were making their selection. [Later on the researchers obtained an 18-inch × 6-inch diameter tube (a "Tinker Toy," tube) closed at one end, into which the beads could be emptied and from which the *S* could then make his drawing. A 6-inch diameter opening is sufficiently restricted to make it impossible for *S*s to see the beads while reaching into the tube.]

The probability of selecting a green bead and winning varies from the certainty (1/1) of winning 10 cents, the one out of two (1/2) chance of winning 20 cents, the 1/3 chance of winning 30 cents, etc., up to the 1/10 chance of winning $1.00. The EV of all the bets is 10 cents so that over a large number of trials none of the bets can be considered more advantageous than any other. The 10 betting cups provide a 10-point scale of risk. This form of the task is enjoyed by subjects and does not appear to present conceptual or instructional difficulties. The measure successfully differentiated between extravert and introvert risk levels among alcoholics who had been given a drink and alcoholics who had not been given a drink (2).

The gambling game (Game I) described above is a somewhat limited measure of risk, since the gambles differ only in the probability of winning varying amounts of money. It is also possible to consider a second type of risk in which gambles differ in the probability of losing varying amounts of money. Finally, it is possible to consider a third type of risk in which the gambles differ in the probability of winning or losing a varying amount of money. In order to assess these other types of risk, the authors have developed Games II and III using the same basic format as in Game I (Appendix 1) above. The three gambling procedures were also designed to be suitable for group administration. The general instructions for the three games were as follows:

NAME: ____________________________

Instructions for Gambling Procedures

In this section of the study you are requested to (*a*) place a bet which may win you some money from the instructor, (*b*) place a bet which may lose you some money to the instructor, and (*c*) place a bet which may either win or lose you some money depending on the toss of a coin.

Each of the three games that you will play offers 10 alternative bets which vary both in the amount it is possible to win (or lose) and in the probability of winning (or losing).

After you have indicated the bet that you wish to make in each of the three games, the outcome of each bet will be determined by having a blindfolded student select one bead from each of the 10 betting cups. If he selects the green bead from the betting cup that you chose for Game No. I, you win the amount of money indicated. If he selects the

> green bead from the betting cup that you selected for Game No. II, you lose the amount of money indicated. If he selects the green bead from the betting cup that you chose for Game No. III, the outcome of a coin toss, heads or tails, determines whether you win or lose.
>
> Inspect the three Game sheets and make sure that you understand the instructions before making your betting cup choices.

As in Game I, specific instructions for Games II and III were put at the head of the game sheets. The subjects were told

> In Game II you hope to avoid losing to the instructor. Circle your preferred betting cup number. Make only one choice.

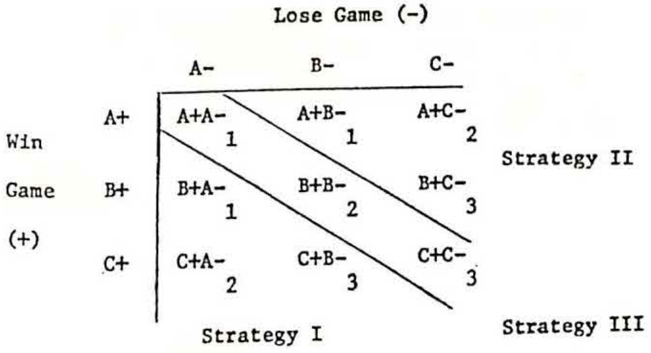
The instructions for Game III stated:

> In Game III you hope to win from and avoid losing to the instructor. Circle your preferred betting cup number. *In addition,* circle the word Heads or the word Tails (directly below these instructions). If the coin toss for your preferred betting cup agrees with your coin toss choice, you win.

The EV for Game I is +10 cents, for Game II, is —10 cents, and for Game III, is zero. The EV summed across all three games is zero, since the wins and losses could be expected to cancel out in the long run. The addition of Game Nos. II and III allows for the collection of more useful information than Game No. I only, and at less cost.

While the responses to the Win Game (I), the Lose Game (II), and the Win or Lose Game (III) can be analyzed as separate scores measuring the players preferred way of coping with the three gambling dilemmas, it is perhaps more useful to consider the independent scores jointly, as a patterning of risk. The authors have developed two approaches for the study of the patterning of risk.

The problems of considering possible patterns of risk taking in the responses of subjects to the games can be facilitated if Game III is disregarded and the 10 risk levels which correspond to the 10 cup numbers are reduced to three. This merely reduces the possible combinations of gambling responses from a possible 1,000 to a more manageable nine, but does not change the basis for the discussion. In Figure 1 the letter A represents the low risk cup numbers 1, 2, and 3; the letter B represents the medium risk cup numbers 4, 5, 6, and 7; and the letter C represents the high risk cup numbers 8, 9, and 10. The plus sign (+) represents the Win Game and the minus sign (—) represents the Lose Game. The subscripts 1, 2, and 3 represent the author's judgment of the three levels of risk within the three Strategies I, II, and III.



A = Low Risk Cup # 1, 2, & 3

B = Medium Risk Cup # 4, 5, 6, & 7

C = High Risk Cup # 8, 9, & 10

1 = Low Risk Level within Strategies

2 = Medium Risk Level within Strategies

3 = High Risk Level within Strategies

FIGURE 1

RISK PATTERNS: STRATEGIES AND RISK LEVELS

Strategy I gambling game choices indicate a preference for a relatively low probability of winning a large amount of money combined with a relatively high probability of losing a small amount. Strategy I emphasizes taking high risks in Game I and low risks in Game II. Within Strategy I, the B+A—$_1$ pattern is least risky because it combines an intermediate level of risk in the Win Game with a low level of risk in the Lose Game and is therefore marked with subscript 1. C+A—$_2$ is medium risky because it combines a high level of risk in the Win Game with a low level of risk in the Lose Game and is therefore marked with subscript 2. C+B—$_3$ has the high risk subscript 3 because it combines a high level of risk in the Win Game with a medium level of risk in the Lose Game.

Strategy II gambling game choices indicate a preference for a relatively high probability of winning a small amount combined with a relatively low probability of losing a large amount. Strategy II involves taking high risks in Game II and low risks in Game I. The meanings of the risk level subscripts for Strategy II are similar to those given for Strategy I.

Strategy III gambling choices indicate a preference for equal probabilities of winning and losing an equal amount of money. A+A—$_1$ is the least risky of all patterns because it has the highest probability of winning or losing the smallest amounts of money. At the other extreme C+C—$_3$ is the riskiest of all patterns because it has the lowest probability of winning or losing the largest amount of money.

The strategies are all equally rational in EV terms, since the EV's of the paired bets within a pattern sum to zero. If one makes the judgment that risk is primarily determined by loss and only secondarily by gain, then Strategy I (where it is possible to win a lot of money but lose little) appears more prudent and less risky than Strategies II and III.

Strategies I and II are perhaps more cognitively complex strategies than III in the sense that they employ one approach for maximizing gain and another approach for minimizing loss, while Strategy III involves a common approach to these two factors.

A second method for deriving patterns of risk was developed as a consequence of a frequency count of *S*'s actual pairings of Win Game and Lose Game cup number choices. The results of this analysis indicated that certain bet choice pairings or combinations occurred far more frequently than others. Appendix 2 depicts a questionnaire (used to study the perception of risk by subjects who

APPENDIX 2

WIN GAME AND LOSE GAME BET COMBINATIONS

Directions

1. What are the bettors trying to accomplish with the bet combinations indicated below? Describe briefly.
2. Please rank the five bet combinations. Put a 1 between the parentheses next to the riskiest combination, a 2 between the parentheses next to the second riskiest bet combination, etc., and a 5 between the parentheses next to the least risky bet combination.
3. *In your judgment* which bet combinations do you consider reasonable and which bets do you consider not reasonable? Reasonable bets are checked yes ______ and not reasonable bets are checked no ______.

Riskiness Rank:		SET B 10¢–$1.00	Reasonable Bet Combinations?
( )	GAME I: WIN GAME	1	Yes ______ No ______
	GAME II: LOSE GAME	1	
( )	GAME I: WIN GAME	1	Yes ______ No ______
	GAME II: LOSE GAME	10	
( )	GAME I: WIN GAME	10	Yes ______ No ______
	GAME II: LOSE GAME	10	
( )	GAME I: WIN GAME	10	Yes ______ No ______
	GAME II: LOSE GAME	1	
( )	GAME I: WIN GAME	5	Yes ______ No ______
	GAME II: LOSE GAME	5	

Additional Comments:

have taken the gambling games) which lists the five most frequently given actual bet choice combinations for the Win Game and the Lose Game. The Bet Combinations represent actual scores which can be subsumed under some of the patterns within the three gambling strategies. The Bet Combinations in research use have the advantage of greater closeness to subjects' actual responses than the Strategies. The Bet Combinations are disadvantageous in that there is considerable attrition of subjects whose responses do not conform to the frequency count-determined patterns. On the other hand, all responses to the Win Game and the Lose Game fit one of the patterns in Figure 1 so that there is no subject attrition when the risk scores are coded in terms of the Strategies.

The data generated by the three gambling games can be analyzed in a variety of ways depending on whether risk taking is considered an independent or dependent variable. The Strategies and the Bet Combinations, the two methods described above for the study of the patterning of risk, involve somewhat qualitative distinctions so that they are perhaps most appropriately used as independent variables. The raw score cup choices for Games I, II, and III are quantitative scales so that they are most appropriately used as dependent variables which can be directly entered into a data analysis in which personality or social characteristics or experimental treatments constitute the independent variable.

The Win Game and Lose Game Bet Combinations (Appendix 2) questionnaire is a useful adjunct to the study of risk considered either as a dependent or an independent variable. The questionnaire, employing a phenomenological approach to the study of risk, frequently clarifies the subjective decision making process followed in responding to the dilemmas of choice in the gambling games.

The most recent games were developed to explore the effects on risk taking of varying the levels of payoff, which in turn modifies the EV. Initial efforts in this direction were restricted to having *S*s make hypothetical gambles in which they were instructed to gamble in all three games as if they were actually playing for real money. When the subjects had completed this task they were told to play three more games, this time for real money.

For the most part subjects respond similarly under the two conditions. A few subjects objected to playing both the hypothetical and real gambling games, while a somewhat larger group were willing to gamble hypothetically but not to gamble for real money. The substitution of the hypothetical payoff games for the real payoff games is not recommended without prior test of possible response differences in the relevant population, since it can be argued that the operational definition of risk taking requires that something of value be at stake.

The Cutter-Heilizer gambling games incorporated the same basic procedures as the earlier games only with an extended range of payoffs (Appendix 3). The instructions are as follows:

Appendix 3

NAME:________________________

CUP NO.	CONTAINS	SET A 1¢–10¢	SET B 10¢–$1.00	SET C 20¢–$2.00	SET D 50¢–$5.00	SET E $1.00–$10.00
1	1 green bead 0 red beads	$ .01	$ .10	$ .20	$ .50	$ 1.00
2	1 green bead 1 red bead	$ .02	$ .20	$ .40	$1.00	$ 2.00
3	1 green bead 2 red beads	$ .03	$ .30	$ .60	$1.50	$ 3.00
4	1 green bead 3 red beads	$ .04	$ .40	$ .80	$2.00	$ 4.00
5	1 green bead 4 red beads	$ .05	$ .50	$1.00	$2.50	$ 5.00
6	1 green bead 5 red beads	$ .06	$ .60	$1.20	$3.00	$ 6.00
7	1 green bead 6 red beads	$ .07	$ .70	$1.40	$3.50	$ 7.00
8	1 green bead 7 red beads	$ .08	$ .80	$1.60	$4.00	$ 8.00
9	1 green bead 8 red beads	$ .09	$ .90	$1.80	$4.50	$ 9.00
10	1 green bead 9 red beads	$ .10	$1.00	$2.00	$5.00	$10.00

	CUP NO.	CUP NO.	CUP NO.	CUP NO.	CUP NO.
GAME I: WIN GAME	____	____	____	____	____
GAME II: LOSE GAME	____	____	____	____	____
GAME III: WIN-or-LOSE GAME	____	____	____	____	____

THE CUTTER-HEILIZER GAMBLING GAMES

Five SETS (A, B, C, D, and E) of three gambling games each are outlined on the other side of the page. The SETS differ only in the range of the AMOUNT OF MONEY that is at stake: SET A varies from 1 cent to 10 cents; SET B varies from 10 cents to $1.00; SET C varies from 20 cents to $2.00; SET D varies from 50 cents to $5.00; and SET E varies from $1.00 to $10.00.

Each SET has three GAMES: GAME I, a WIN game, GAME II, a LOSE game, and GAME III, a game that may turn out to be either a WIN game or a LOSE game, depending upon the subsequent toss of a

> coin. In GAME I, the WIN game, the selection of a green bead from a cup results in a gain of the amount of the bet. (There is no gain or loss if the green bead is not picked.) In GAME II, the LOSE game, the selection of a green bead from a cup results in a loss of the amount of the bet. (There is no loss or gain if the green bead is not picked.) In GAME III, the WIN-OR-LOSE game, the selection of a green bead from a cup results in a gain or a loss of the amount of the bet. A subsequent toss of a coin will determine which it is to be—a gain or a loss. (There is neither a gain nor a loss if the green bead is not picked.)
>
> Ten CUPS (1, 2, 3, 4, 5, 6, 7, 8, 9, and 10) are used in each of the GAMES of each of the SETS. Each CUP contains ONE GREEN BEAD and a number of RED BEADS that VARIES from NO RED BEADS in CUP 1, ONE RED BEAD in CUP 2, TWO RED BEADS in CUP 3, and so on with ONE RED BEAD being added to each succeeding CUP until there are NINE RED BEADS in CUP 10. In other words, the TOTAL NUMBER OF BEADS in each CUP corresponds to the NUMBER of the CUP, and there is always ONE GREEN BEAD in each CUP.
>
> IF YOU WANT TO BET, YOU MUST BET ON ALL THREE GAMES IN A SET. YOU CANNOT BET ON ONE OR TWO OF THE GAMES IN A SET. YOU MAY BET ON ANY NUMBER OF SETS, BUT MUST ALWAYS BET ON A LOWER SET IN ORDER TO BET ON A HIGHER SET. (For example, SET A must be bet in order for SET B to be bet, and so on.)
>
> The three bets for each SET are to be indicated in the appropriate spaces underneath each SET by writing in the NUMBER of the CUP that you want to bet: writing in the CUP NUMBER for GAME I, the CUP NUMBER for GAME II, and the CUP NUMBER for GAME III. After all bets have been placed, and the gambling choices have been collected, volunteers will blindly select a BEAD from each CUP. It is these selections that will determine whether a cup WINS, LOSES, or WIN-LOSES. After GAME III, the toss of a coin will determine whether the GREEN BEAD selections for GAME III represent GAINS or LOSSES.
>
> Don't forget to write your name in the appropriate place on the other side.

In the Cutter-Heilizer games Set B corresponds to the level of payoffs employed in the earlier games. Set A extends the range of payoffs downward and Sets C, D, and E extend the range payments upward.

There are several factors to consider when using the Cutter-Heilizer games. First of all, the games are potentially expensive for the *S*s or for the researcher, since it is possible that each *S* could conceivably win or lose a maximum of $36.20. This outcome is highly unlikely if the 150 possible bead drawings are used. The analysis of the data generated by the additional game sets differs from that described earlier because the relatively high potential gains and losses cause many subjects to stop playing before completing all the sets. At the moment the authors have restricted themselves to using the simple *number* of sets played as an additional index of risk. For those subjects who go on to gamble in Sets C, D, and E, it would be of particular interest to study the frequency of shifts from a strategy that has potential for a large loss and little gain at the lower Sets A and B to a strategy that emphasizes large gain and little loss. The high loss levels of Strategies II and III become potentially more punishing at the higher payoff levels. Strategy I should become increasingly attractive to the prudent gambler at the higher levels of payoff.

The experience with the present gambling games suggests some further developments that would enhance their power in future research. It is apparent that the average college student does not normally think mathematically and is usually unable to subject the alternative bets to rigorous analysis. Many of the subjects required a lengthy question and answer period before they understood the games' characteristics. If performance in the games (which presumably measures the tendency to take gambling risks) is confounded with the ability to solve problems on a novel task, then many of the personality characteristics that have been related to risky behavior (3) would disappear if the game format were simplified, the probabilities were computed, and the EV of the bets was made more intuitively obvious and more explicit by instructions. For example, instead of merely giving the *S*s the numbers of red and green beads in the gambling cups, the payoffs, and letting them figure out the probabilities and expected values from this raw data, one might state the probability level of each cup as a percentage in the same manner that modern day weather men refer to a 10 per cent chance of rain. The subject might also find it easier to think about gambling odds if the total number of beads was held constant at 100 with only the proportion of red and green beads varying. Thus the 10 per cent chance to win a dollar would be substituted for the 1/10 chance to win a dollar provided by Cup No. 10 (Set B) which contains one green bead and nine red beads. A cup representing a 10 per cent chance would contain 100 beads: 10 green, and 90 red. As the following list of suggested bets indicates, this format would allow an extension of the range of probabilities down beyond the 1/10 chance level limit to a minimum 1/100 chance allowing a more sensitive measure of risk. This new method would also reduce the number of Sets from five to two, the number of alternative bets from 10 to seven, and at the same time maintain the range of payoffs from 1 cent to $10.00.

Cup A contains 100 green beads and, since green is the "action" color, the subject can always expect to Win (or Lose) 1 cent or 10 cents each time he chooses to gamble from this cup. Cup B contains 50 green beads and 50 red beads so that one half of the time the gambler can expect to select a green bead. It can be seen that the percentage chance of winning or losing is directly represented by the exact number of green beads in each cup (Appendix 4).

APPENDIX 4

SIMPLIFIED CUTTER-HEILIZER GAMBLING GAMES BETS

Cup*	Chance to Win (or Lose): (number of green beads in cup)	Set A	Set B
A	100%	1¢	10¢
B	50%	2¢	20¢
C	20%	5¢	50¢
D	10%	10¢	$ 1.00
E	5%	20¢	$ 2.00
F	2%	50¢	$ 5.00
G	1%	$1.00	$10.00

* 100 beads in each cup.

To make sure that the subjects recognize that the EV's are held constant, the instructions could point out that the percentage chances of winning and the amounts to be won are balanced so that the greater chances to win are correspondingly matched by smaller potential winnings.

The *S*'s decision problem could be simplified with a net gain in the amount of information obtained if a paired comparisons technique was used instead of choosing a single bet from among the alternatives. This procedure could be followed separately for the Win Game, the Lose Game, and the Win or Lose Game. The subjects would also be required to play both Sets of all three games if they choose to play at all, since the potential losses could be limited to 11 cents for any subject.

C. SUMMARY

In summary, it can be seen that the procedures discussed above have developed from a fairly simple game in which subjects have an opportunity to win a small amount of money to a complex game in which they can win or lose a fairly large quantity of money. The uses of Strategies and Bet Combinations were discussed in terms of their utility in employing risk as an independent variable. Preliminary analysis of subjects' perception of popular Bet Combination responses in the Win Game and the Lose Game suggests that the perception of risk by the subjects is an important variable to consider in future research. It was suggested that the apparent difficulty that some subjects have in performing the games may reflect a cognitive or personality variable confounded with that of risk, which could be eliminated by various procedures that would simplify information processing preceding gambling bet choice. The authors hope that their experience in developing a series of research gambling games will be of use to others and will stimulate the development of gambling games by others.

REFERENCES

1. COHEN, J. Chance, Skill, and Luck. Baltimore: Penguin, 1960.
2. CUTTER, H. S. G., HARFORD, T. C., & GREEN, L. R. Levels of risk taken by extraverted and introverted alcoholics as a function of drinking whiskey. Paper presented at the convention of the American Psychological Association, New York, September, 1966.
3. KOGAN, N., & WALLACH, M. A. Risk Taking, A Study in Cognition and Personality. New York: Holt, Rinehart & Winston, 1964.

Veterans Administration Hospital
Brockton, Massachusetts 02401

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 33-38.

RATIONALITY, MOTIVATION, AND EXTINCTION*[1]

Department of Psychology, Oklahoma State University

ROY GLADSTONE AND MERCEDES MILLER

A. INTRODUCTION

A fundamental function of a science is to enable accurate prediction and control. Since rationality can be used to produce unique answers to some kinds of problems, it follows that it ought to be possible to eliminate error in predicting individual behavior if (*a*) *S* can detect a unique solution, (*b*) we present a problem with a unique solution, and (*c*) *S* uses that solution. Obviously, this would represent a considerable improvement over the probabilistic predictions inherent in the use of current laws or principles of psychology. Translating this theoretical hunch into an experiment involves finding a problem that can be solved both rationally and in some other way. The problem chosen was that faced by an *S* when extinction starts in operant conditioning—should he stop responding or not? When the problem is susceptible to a rational solution, it is possible to hypothesize that every *S* will stop responding at the beginning of the extinction period: i.e., act rationally.

B. RELATED LITERATURE

There are four pertinent ways of explaining extinction phenomena. Ferster and Skinner (2) explain them in terms of schedules of reinforcement. Perception and significance are totally ignored.

Bitterman *et al.* (1) explain extinction in a second way by hypothesizing that "the rate of extinction is inversely related to the similarity between conditions of training and extinction" (1, p. 456). They go on to show that their results can be explained by this "discrimination hypothesis," while the concept of secondary reinforcement cannot explain them.

Gladstone (4) pitted the Skinnerian explanation against that of Bitterman. Keeping the schedule of reinforcement the same for two groups, a light was

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 10, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The authors are most grateful for the aid which made this project possible. The experiment is an outgrowth of another carried out with the aid of a Cooperative Research Project Grant from the U.S. Office of Education. The Research Foundation of the Oklahoma State University has underwritten the costs of publication. Dr. Thaddeus Cowan was kind enough to assemble the equipment.

turned on at the beginning of the extinction period for group 1 but not for group 2, thus reducing the similarity between the training and the extinction conditions for group 1 more than for group 2. Under these conditions the extinction responses of group 1 were reduced considerably below those of group 2.[2]

A third way of explaining the phenomena is through the use of the Tolmanian construct of expectancy. The hypothesis is that *S* will not yield extinction responses (Rs) if he does not expect the reward to result. The nonresponse extinction experiment was devised to test this hypothesis. The literature on nonresponse extinction is reviewed by Kimble (5) and commented upon by Gladstone (3). Kimble concludes that nonresponse extinction does occur. Gladstone's (3) results are consonant with that conclusion. As hypothesized, Gladstone found that when college students were able to see the rewards empty out of a reservoir and the resultant empty reservoir, their behavior in the extinction situation differed from that of students who could not see the rewards. However, the behavior of the *S*s who could see the rewards could not be described with assurance as rational; while 32 per cent (unpublished) of the *S*s stopped abruptly when the last reward fell from the reservoir (i.e., yielded nonresponse extinction), 68 per cent did not. Gladstone also found that sex made no significant difference in these results.

Gladstone (4) pitted the Bitterman similarity hypothesis against the Tolmanian expectancy hypothesis, but the results were ambiguous. It is possible to attribute the ambiguity to insensitivity due to the experimental design used.

A fourth explanation bearing upon extinction phenomena is that of motivation. Hypothesizing that the apparently irrational behavior of 68 per cent of the *S*s in Gladstone (3) could be controlled by giving the *S*s a reason for *not* giving extinction Rs, Gladstone (4) essentially repeated the earlier procedure but took money away from *S* for each extinction R in one group but not in the other. Both groups were able to see the emptied reward reservoir. In the group that had no penalty for giving extinction Rs, 27 per cent of the *S*s stopped abruptly when the reward reservoir emptied, a percentage very close to the 32 per cent of Gladstone (3). In the group that had a penalty for giving extinction Rs, 70 per cent stopped abruptly, a considerable improvement in accuracy of prediction. However, the latter results were obtained from an *N* of 10.

---

[2] A parametric test of the Bitterman hypothesis is projected.

C. RATIONALE AND DESIGN

The goal was the complete elimination of extinction Rs: i.e., complete control over the behavior of every *S*. The four ways of explaining extinction behavior described above suggest different approaches to this end. The schedule of reinforcement approach suggests a continuous schedule of reinforcement but offers no hope of gaining complete control. The Bitterman approach suggests that the control desired increases with an increase in difference between the training and extinction situations, but does not suggest that the characteristics of change are important. The Tolmanian approach suggests that the significance of the cues available is important. Within the context of this experiment the nature of significance is quite clear, although this might not be the case in other situations. Finally, the motivation approach suggests that motivation-to-stop is important.

This experiment combines the motivation approach with the Bitterman approach on the one hand and the Tolmanian approach on the other and pits one resultant situation against the other, the question being which will give most precise control for the desired end.

1. *Equipment*

The equipment consisted of (*a*) a rat pellet feeder with 10 BBs substituted for pellets, (*b*) a knob which was pulled in order to operate the feeder, (*c*) a programer which fed the BBs on a variable ratio reinforcement schedule which was the same for each *S*, and (*d*) a light. The feeder had two covers, one opaque and the other transparent, so that the rewards were either visible or invisible depending on the cover used. The light could be set to go on when the last BB was released. *S*s were paid for each BB as described below. Extinction Rs were recorded mechanically.

2. *Conditions*

There were two conditions. One was the normal extinction condition in which *S* could not tell when extinction started, changed only by the fact that a light went on in *S*'s face when the last reward dropped from the reservoir. The second was one in which the BBs (token rewards) and the exhaustion of the reward reservoir at the end of the reward period were both visible. These conditions are referred to as the invisible-with-light and visible conditions, respectively. Contrasting the visible condition with the invisible-with-light condition enables a test of the effect of the perception of emptiness and of the rational conclusion which that perception makes possible against the effect of simple amount of change.

3. *Subject-Equipment Interaction*

The nature of the equipment was such that it was difficult to stop responding at the time the last BB fell. The effect of this difficulty was assessed in Gladstone (4) by requesting 30 *S*s to respond at a normal speed for them but to stop as soon as the last BB fell. The average number of extinction Rs was 1.2. Four *S*s yielded four extinction Rs with three yielding three. This means that even when the typical *S* intended to stop at a given point he was unable to do so. This can be interpreted, if desired, in terms of physiological response time or in terms of the fact that it took time for the BB to fall and strike the dish with a clatter thus completing the entire signal.

4. *Procedure*

*S*s signed up for a specific time. At the time a given *S* arrived at the experimental room the apparatus was set for a given stimulus condition. The next *S* found the second stimulus condition, etc.

5. *Directions*

The following directions were used.

> *Procedures.* The subjects in each group will be taken one at a time to the experimental room where they will stand in front of the apparatus and be given the following instructions:
>
> "There are no tricks in this experiment. Everything is just as it appears to be. We are trying to see how far down in the age scale we must go before people start to act in a way which doesn't make any sense. We expect you to act sensibly. Just follow the directions in a way which makes sense to you.
>
> "Your task in this experiment will be to operate this machine. Here is how it works. Pull this plunger out and push it in and a small BB will drop into this cup, like this (*E* demonstrates). Later you will be given two cents[3] for every BB you have. You will have 1/5¢ taken away for each time you pull the plunger. Do you understand what you are to do?" (*E* may repeat the essential instructions but questions as to the nature of the experiment will be answered pleasantly with the phrase, "I am not allowed to tell you any more about the experiment. Just act as sensibly as you can.")
>
> "You may start now. Please tell me when you are through." (*E* responds to *S*'s question, "Can I stop now?" with "It's up to you.") When *S* says he is through, *E* will give the reward to *S* in exchange for the number of BBs in the reward cup. If *S* obviously has stopped but does not say so, say, "Are you through?" If *S* indicates he is, give him the reward.

---

[3] The reward in Gladstone (4) was 15¢, the penalty being 1¢.

E asks *S* why he stopped responding and records *S*'s answer. Say, "Please do not discuss this experiment with anyone else."

6. *Subjects*

Two groups of 30 college students each were used. One group was subjected to the visible condition, the other to the invisible-with-light condition.

7. *Analysis Rationale*

A corollary of the expectation of rational control of extinction behavior is the expectation that a significant number of scores will pile up near zero: i.e., the distribution will not be normal. Similarly, it is anticipated that the variances of the two groups of scores will be very different. However, scores can be dichotomized into rationally controlled *vs.* nonrationally controlled. Given this character of the data, chi square was used for analysis. In the light of the subject-equipment interaction described above, an extinction score of two or less was used as an indication of rational control.

D. Results and Analysis

Because exact scores are of more interest than means, the entire distribution is given in Table 1. Chi square is about 19 where 6.6 indicates a difference significant at the .01 level. The number of rationally controlled extinction scores in the visible, motivated condition is 73 per cent, corresponding closely with the 70 per cent figure of Gladstone (4) despite the change in amount of reward and penalty.

TABLE 1

Number of Extinction Responses for Each *S*

Invisible-with-Light			Visible		
15	9	5	1	0	3
55	30	8	0	6	4
10	190	10	0	24	1
29	5	59	0	42	1
7	35	1	2	2	11
18	2	5	1	0	0
15	160	17	1	0	1
19	13	31	0	10	1
0	22	22	0	1	0
16	0	1	1	3	0

E. Conclusions

Extinction behavior in the classical Skinnerian experiment is controlled largely by the schedule of reinforcement. When signals are added at the end of the reward period, the mean number of extinction Rs is reduced. This is true whether the signal is arbitrary, as is a light, or whether the signal is rationally interpretable, as is the visible lack of rewards.

Best control, although not complete control, was produced by combining the rational cue with motivation to stop. This dual mode of control enables precise prediction of the behavior of a large percentage of *S*s as well as better control over the mean. The degree of precision is very much greater than is possible through the use of the typical principle of psychology.

Thus it was useful to theorize that college students are goal-directed and are capable of using cues rationally. It may be possible to interpret these results within a Skinnerian framework (6), but a Tolmanian position seems much more comfortable for the purpose.

F. Summary

The goal *E* set for himself was the complete elimination (control) of extinction responses. Two groups were used. In both groups a reward gained during training was taken away for each extinction response. In one group *S* could see the reward reservoir empty out during the training period. In the other group the rewards were hidden but a bulb was lit in *S*'s face when the last reward fell from the reservoir. The first situation produced greater control than the second situation, but complete control was not achieved.

References

1. Bitterman, M. E., Fedderson, W. E., & Tyler, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456-464.
2. Ferster, C. B., & Skinner, B. F. Schedules of Reinforcement. New York: Appleton-Century-Crofts, 1957.
3. Gladstone, R. The significance of the extinction situation. *J. of Psychol.*, 1966, **64**, 23-28.
4. ———. Effect of extrinsic rewards on motivation. Cooperative Research Project No. 5-8164-2-12-1, U.S. Dept. of Health, Education and Welfare, 1966.
5. Kimble, G. A. Hilgard and Marquis' Conditioning and Learning (2nd ed.). New York: Appleton-Century-Crofts, 1961.
6. Skinner, B. F. Operant behavior. In W. K. Honig (Ed.), *Operant Behavior: Areas of Research and Application.* New York: Appleton-Century-Crofts, 1966.

Department of Psychology
Oklahoma State University
Stillwater, Oklahoma 74074

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 39-48.

INDUCTION OF ACTION AND ATTITUDE CHANGE: THE FUNCTION OF ROLE-SELF CONFLICTS AND LEVELS OF ENDORSEMENT*[1]

Veterans Administration Hospital, Brockton, Massachusetts

PHYLLIS G. ORAM

A. INTRODUCTION

A number of studies have shown that when a subject is induced to behave in a manner opposed to his private opinion or attitude, his attitude tends to change to conform more closely to the induced behavior. One explanation for this change is that it is the result of improvisation. According to this view a person who accepts the task of role playing or improvising becomes temporarily motivated in favor of the new position, and is subject to self-persuasion. This position is called incentive theory (3, 5, 6, 8).

Cognitive dissonance is also advanced as an explanation for change in attitude under induced behavior conditions. Brehm and Cohen (2), in their amplification of Festinger's theory, state that induced behavior leads to dissonance, which has motivating characteristics. Since the cognition arising from the induced behavior is relatively resistant to change, the cognition related to attitude is most likely to change to conform with behavior and reduce dissonance.

Kelman (7) suggests that attitude change may occur as a result of induced behavior for any one of a variety of reasons. One of the major factors influencing whether or not attitude change will occur is the relationship that the person has to the attitude object before the induced behavior. This relationship of the person to the attitude object can be defined in terms of approach and avoidance tendencies. The notion of approach and avoidance tendencies as differential predictors of behavior derives from the goal gradient model or conflict model developed by Miller (9, 10).

In the goal gradient model, when drives are based on extrinsic cues response tendencies will be at their height close to the goal and will fall off sharply as distance from the goal increases. When drives are based on intrinsic cues the

---

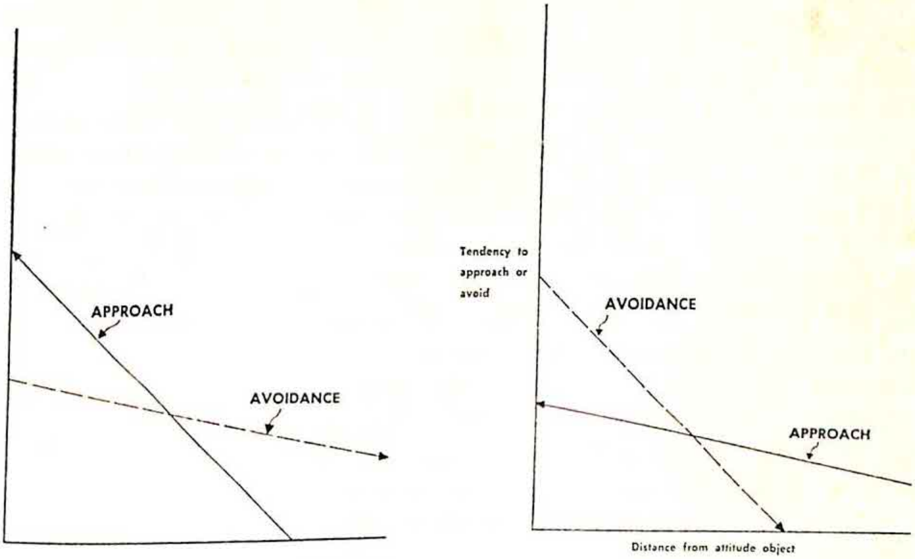
* Received in the Editorial Office, Provincetown, Massachusetts, on October 13, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This article was adapted from a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree at Boston University.

specific goal is not as salient in maintaining drive level, and response strength will be relatively equal at all points of distance from the goal. While the original Miller studies postulated a steep avoidance gradient and a shallow approach gradient, Miller's later (10) work has made the more general statement that approach *or* avoidance tendencies based on motivation elicited by external cues will vary sharply as a function of distance from the goal, while approach or avoidance tendencies primarily dependent on internal cues will have shallower gradients of generalization. Induced behavior will lead to attitude change where there is a steep *approach* gradient. Induced behavior will not lead to attitude change where there is a steep *avoidance* gradient.

Using the goal gradient model, Kelman (7) describes attitude change in situations where response tendencies are related to identification and internalization. The present study examines situations with conflicting response tendencies arising from components of role and self. The goal gradient relationships of role and self are similar to those of Kelman's identification and internalization terms. That is, the behaviors and attitudes associated with a role are most salient when the person is in the role, and become much less salient when the person moves away from that role. For this reason role, in this study, is considered as having the properties of extrinsic motivation. Thus the approach or avoidance tendencies associated with a role would have steep gradients of generalization, at their height close to the goal and falling off sharply at a distance from the goal. On the other hand, the behaviors and attitudes associated with the central self operate over a wide variety of situations, and for this reason self is considered to have the properties of intrinsic motivation. Relationship to a particular goal is not required for maintenance of response strength, and thus the gradient of generalization is shallow, with response strength relatively equal at all distances from the goal. This definition of self as intrinsic motivation conforms to the work of Heilizer (4). He has applied the conflict model to social motivation, and suggests that social behaviors that have been consistently rewarded over a long period of time by many people in a wide variety of situations can be considered as intrinsic motivation. On the other hand, Berkowitz, Malone, and Klein (1), who have also dealt with role concepts in terms of goal gradients, regard role behavior as internalized and thus associated with intrinsic motivation. In the present study behaviors associated with role are seen as primarily dependent on situational cues and thus contain the properties of extrinsic motivation. This view conforms with Kelman's position.

Figure 1 illustrates the two situations of interest in this study: A steep approach—shallow avoidance situation, and a shallow approach—steep avoidance situation. Since we are interested in steep and shallow gradients of role



A. Role-approach/self-avoidance

B. Self-approach/role-avoidance

FIGURE 1
ILLUSTRATION OF THE TWO GOAL GRADIENT SITUATIONS

and self, these two situations will be referred to as role-approach—self-avoidance and self-approach—role-avoidance.

If a subject is induced to behave in a manner opposed to his private opinion when his relationship to the attitude object (that toward which he is behaving) is based on one of the two types of conflict situations described above, the following hypotheses regarding attitude change can be derived:

1. (*a*) If the relationship is a role-approach—self-avoidance (steep approach—shallow avoidance) association, induced behavior will lead to attitude change in the direction of the induction. (*b*) Induced behavior which requires the subject to endorse the attitude object strongly will lead to more attitude change than induced behavior which does not require strong endorsement.

2. If the relationship is a self-approach—role-avoidance (shallow approach—steep avoidance) association, induced behavior will not lead to attitude change.

B. METHOD

The role of the professional nurse was used to examine these situations. Subjects were 87 registered nurses who had returned to college to obtain the B.S. degree.[2] The experimental sessions were held during class time.

---

[2] The author wishes to thank Myrtle Ford, R.N., and Marita Tribou, R.N., Boston University School of Nursing; and Mary Welter, R.N., Boston College School of Nursing, for permission to use their classes in this experiment.

Each class participated in two experimental sessions. In the first session background data were collected, two conflict questionnaires used for assigning subjects to experimental conditions were given, and the attitude scale was administered for the first time. At the second session approximately two weeks later, subjects wrote the essay and the attitude scale was administered again. All scales, questionnaires, and essays were anonymous (with the one exception noted below), and the data for each subject were identified only by number.

Each subject was asked to write an essay against her own opinion on one of three conflict areas. The topic assigned was considered to be a conflict area for that particular subject on the basis of the questionnaires administered in the first experimental session. Two of the topics were based on role-approach—self-avoidance situations, and one topic on a self-approach—role-avoidance situation. A second self-approach—role-avoidance topic was originally included as part of the study, but the attitude scales associated with the topic did not prove to be reliable, and the topic was dropped.

The two topics based on role-approach—self-avoidance situations were as follows:

Evaluation Conflict. Approach: Role requirement for evaluation and supervision of auxiliary nursing personnel; *Avoidance:* Personal dislike for behaviors required in evaluation and supervision.

Morality Conflict. Approach: Role requirement of treating all patients alike, including those with "immoral" illnesses; *Avoidance:* Personal dislike for associating with "immoral" persons.

The topic based on a self-approach—role-avoidance situation was as follows:

Involvement Conflict. Approach: Tendency to become personally involved with the patients or nurses; *Avoidance:* Professional disapproval of personal involvement with patients.

The three topics above were chosen as being relatively common conflicts in nursing, based on a search of nursing literature and a questionnaire to 18 experienced nursing supervisors.

The first experimental condition was the topic of the essay, whether it was a role-approach topic or a self-approach topic. The second experimental condition was the level of endorsement required in writing the essay. Subjects were randomly assigned to one of three endorsement conditions. The three conditions were (*a*) *High Endorsement:* The subject was told to give arguments in favor of the position, to sign her name to the essay, and was told that a panel of nursing supervisors in another city would see her essay; (*b*) *Low Endorsement:* The subject was simply asked to give arguments in favor of the position; and (*c*) *Neutral:* The subject was asked to describe an incident related to the topic rather than give arguments.

Instructions as to topic and endorsement conditions were given each subject as the cover sheet of a booklet in which the essay was written. In the High and Low endorsement conditions, the instruction sheet also included three possible arguments which the subject could use as part of her essay if she chose.

The dependent variable in this experiment was change in attitude scale scores from pre-essay writing to postessay writing. The attitude scale contained six Likert-style items related to each of the three subject areas. Test-retest correlations for each of the subscales were significant at $p < .01$ (Evaluation $= .89$, Morality $= .79$, Involvement $= .82$). A typical item from the scale is presented in Figure 2.

"Evaluations and criticism of auxiliary personnel are the most unpleasant part of my job."

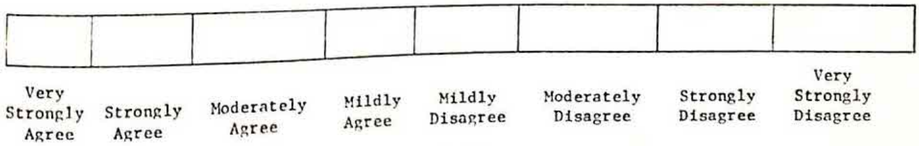


FIGURE 2
A TYPICAL ITEM FROM THE ATTITUDE SCALE

C. RESULTS

Table 1 shows the analysis of variance of the difference scores for the three experimental groups on the relevant subsection of the attitude scale. The analysis shows that there is a significant difference between pre- and postexperimental attitude scores at $p < .01$ for the main effect of type of conflict situation, but no difference for the main effect of endorsement or for the conflict situation by level of endorsement interaction.

TABLE 1
ANALYSIS OF VARIANCE: DIFFERENCE SCORES FOR THREE EXPERIMENTAL GROUPS
($N = 87$)

Source of variation	SS	df	MS	f
Conflict situation	382.241	2	191.12	5.68**
Endorsement	206.096	2	103.05	3.06
C × E	194.423	4	48.61	1.45
Error	2624.056	78	33.64	

** $p < .01$.

Examination of the mean difference scores for each of the three conflict situations shown in Table 2 indicates that there was some tendency in all three groups to shift toward the self end of the attitude scale after the induction, since a minus score indicates a shift toward self items. However, t tests of the differences between the means show the shift for the role-approach group Evaluation over the self-approach group Involvement to be significant at $p < .01$, while the difference for the role-approach group Morality over the self-approach group Involvement is significant at $p < .10, > .05$. The difference between the two role-approach groups is not significant. These findings generally support the predictions that induced behavior would lead to attitude change in role-approach situations but not in the self-approach situation.

TABLE 2
MEAN DIFFERENCE SCORES BY CONFLICT SITUATION

Type of situation	Mean difference
Role-approach	
Evaluation	—5.31
Morality	—3.22
Self-approach	
Involvement	— .55

While the findings support the prediction for type of conflict situation, the prediction relative to the second major treatment condition—level of endorsement—has not been supported. It was predicted that strong endorsement would lead to more attitude change than weak endorsement in the role-approach situations, but Table 1 shows no difference in amount of attitude change between the three endorsement conditions in either of the role-approach situations. An explanation of this lack of difference between endorsement conditions may perhaps be found if we examine the actual induction, essay writing. Table 3 summarizes the results of a content analysis of these essays. The content analysis was carried out by two judges independently. A reliability coefficient of .79 was obtained between the two judges.

In the High endorsement and Low endorsement conditions, directions to the subjects asked them to give arguments in favor of the stated position. In general the subjects followed directions and produced the kinds of arguments they were asked to give. In most cases they also produced a few arguments in the other direction as well, but the ratio of these anti arguments to pro arguments was quite low. Production of arguments *opposed* to the stated position is not unusual, and will be found, for example, in Janis and Gilmore (5).

The instructions to the subjects in the Neutral condition asked them to describe an incident related to the topic, or to discuss it generally. But although these subjects were not instructed to give arguments pro or con, they did so. And since they were not under instructions to produce arguments limited to one side of the position, they tended to produce both pro and con arguments. In all three groups subjects writing under the neutral conditions produced more arguments against the topic than did subjects writing under the other two conditions. The instructions given to the subjects seem to have produced a somewhat different type of induction situation from that produced by the instructions given under the High and Low endorsement conditions, and it is dubious whether this condition can be called "neutral" as that term is usually understood.

D. Discussion

This study gives limited support to the notion that attitude change based on improvisation can be predicted on the basis of the approach and avoidance gradients inherent in the situation. In those situations where there is a steep approach gradient (role-approach situations), production of arguments against one's own position leads to attitude change. In the situation with a steep avoidance gradient (self-approach situation), improvisation did not lead to attitude change. These results support predictions based on the goal gradient model as opposed to either Incentive Theory or Dissonance Theory. Both Incentive and Dissonance theories would predict attitude change in all three of the conflict situations tested in this experiment. Only the goal gradient model makes a distinction between role-approach—self-avoidance situations and self-approach—role-avoidance situations. It is this distinction that leads to the prediction of no attitude change under the self-approach—role-avoidance situation.

There was no difference between High and Low endorsement conditions in any of the groups. Since a number of studies have found differences in amount of attitude change as a result of differences in the induced behavior, the lack of difference between these two treatments in the present study requires some explanation. Perhaps the explanation can be found in the similarities among the levels of endorsement, rather than in their differences. In their 1954 and 1956 studies, Janis and King found differences in amount of attitude change when subjects performed different tasks: that is, when they actually role-played *vs.* read a communication or listened to a speech. Janis and Gilmore (5) found differences in amount of change between subjects who wrote an essay and those who agreed to write but did not actually do so. In neither case, however, were there variations of the actual conditions of writing the essay. In the present experiment both High and Low endorsement groups wrote essays, supposedly under varying conditions. The fact that no differences were found between them suggests that *actually writing the essay* may be the important ingredient in inducing attitude change.

Further evidence of the influence of the specific task can be found when we compare all three endorsement conditions. The assumption was made that the Neutral groups would act as control groups and that, since they did not give arguments, there would be no attitude change. However, the subjects did in fact give arguments, both for and against the position. These essays might be said to be a review of the topic, rather than a one-sided presentation. Under these conditions in both role-approach groups there was attitude change under the Neutral condition. This change in attitude after a review of all arguments supports Scott's (11) position in opposition to dissonance theory that dissonant cognitions need not be motivating unless the holder is confronted with them and the task of defending or explaining them. We might speculate that it is impossible to take a neutral position on a topic that represents a personal conflict area—and these topics were chosen to be personal conflicts. Perhaps *any* discussion of a conflict topic, however neutral or balanced, has motivating characteristics. This possibility emphasizes the difficulties and the importance of developing a true Neutral condition for control purposes.

The fact of attitude change in the Neutral conditions brings up the question of the adequacy of the control groups in this study. However, another partial control group is available: the reliability sample used to test the attitude scale originally. The sample consisted of 25 students in the same program at Boston University as the students who took part in the experiment. The high test-retest reliability coefficients obtained from these students, as well as the fact that there were no significant changes in the means of the distributions, is some evidence that the attitude change found in the experimental sample can be attributed to the induction.

An additional point of interest is raised by the content analysis of the essays. Zimbardo (12) and others have questioned the use of the term improvisation in the Janis and King experiments, and have pointed out that there are few, if any, evidences of the creation of new arguments beyond those supplied to the subject in the original communication. Indeed, the Janis and Gilmore (5) study indicates an average of less than one pro argument per subject. In the present study the mean number of pro arguments offered was 1.77, and the mean number of new pro arguments above and beyond those supplied to the subjects in the original communication was .55, which is slightly more than previous studies have obtained. We suggest that the difference arises because in this experiment improvisation was around an important professional role. It was a topic close to the subject, and one on which she was prepared to talk. Thus the overall impact of the essay writing task might be different from what it would be in a situation wherein the subject was asked to think up arguments about a topic with little relevance to his day-to-day life.

E. SUMMARY

This study explored the conditions under which induced behavior leads to attitude change. Two types of situations were investigated, a role-approach—self-avoidance situation, under which it was predicted attitudes would shift toward the induced behavior, and a self-approach—role-avoidance situation in which no change was predicted. It was also predicted that more attitude change would occur under conditions of strong endorsement than under conditions of weak endorsement. In the experiment nurses were asked to write essays against their own opinions under one of the two types of conditions. The results support the notion that attitude change can be predicted on the basis of the goal gradients inherent in the situation.

REFERENCES

1. BERKOWITZ, N., MALONE, M., & KLEIN, M. Simultaneous role conflict, compromise and tension. Paper read at Eastern Psychological Association, New York, New York, 1966.
2. BREHM, J. W., & COHEN, A. R. Explorations in Cognitive Dissonance. New York: Wiley, 1962.
3. ELMS, A. C., & JANIS, I. L. Counter-norm attitudes induced by consonant versus dissonant conditions of role-playing. *J. Exper. Res. Personal.*, 1965, **1**, 50-60.
4. HEILIZER, F. Conflict models, alcohol, and drinking patterns. *J. of Psychol.*, 1964, **57**, 457-473.
5. JANIS, I. L., & GILMORE, J. B. The influence of incentive conditions on the success of role playing in modifying attitudes. *J. Personal. & Soc. Psychol.*, 1965, **1**, 17-27.
6. JANIS, I. L., & KING, B. T. The influence of role-playing on opinion change. *J. Abn. & Soc. Psychol.*, 1954, **49**, 211-218.
7. KELMAN, H. C. The induction of action and attitude change. In S. Coopersmith (Ed.), *Personality Research*. Copenhagen: Munksgaard, 1962. Pp. 81-110.
8. KING, B. T., & JANIS, I. L. Comparison of the effectiveness of improvised vs. nonimprovised role-playing in producing opinion changes. *Hum. Rel.*, 1956, **9**, 177-186.
9. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the Behavior Disorders* (Vol. 1). New York: Ronald Press, 1944. Pp. 431-465.
10. ————. Liberalization of basic S-R concepts: Extension to conflict behavior, motivation and social learning. In S. Koch (Ed.), *Psychology: A Study of a Science* (Vol. 2). General systematic formulations, learning and special processes. New York: McGraw-Hill, 1959. Pp. 196-292.
11. SCOTT, W. A. Cognitive consistency, response reinforcement and attitude change. *Sociometry*, 1959, **22**, 219-229.
12. ZIMBARDO, P. G. The effect of effort and improvisation on self-persuasion produced by role-playing. *J. Exper. & Soc. Psychol.*, 1965, **1**, 103-120.

Veterans Administration Hospital
Brockton, Massachusetts 02401

Published as a separate and in *The Journal of Psychology,* 1968, **68**, 49-53.

READING EASE AS A FACTOR IN IMPROVED COMMUNICATION EFFECTIVENESS*[1]

Pacific State Hospital, Pomona, California; and Pomona College

CHARLES V. KEERAN, JR., AND GRAHAM B. BELL

A. INTRODUCTION

Much of the formal communication system in a large organization relies upon written messages. Administrators, scholars, and employees have all expressed the belief that the complicated style of such messages frequently reduces their communication effectiveness. Descriptive studies underscore this belief. Bell and Hoffman (1) concluded that "one serious problem is that many interoffice memos are poorly written and consequently ignored." Mathies (4) reported the frequent complaint that "procedure manuals are too wordy; there are too many references to other procedures; they are too hard to understand."

In recognition of these problems, training courses have been developed and articles have been written on techniques of more effective writing. Yet how much improvement in communication effectiveness can be expected even if messages are better written? The present paper reports on an empirical study addressed to this question.

B. METHOD

The study was conducted in an operating organization—a large state hospital. The basic design was to distribute a message prepared in two different styles, with the content held constant. The communication effectiveness of the two styles was then compared.

1. *Experimental Message*

The research was carried out without modifying or disrupting the usual flow of work. The fire rules of the hospital were scheduled for revision, and

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 16, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Supported in part by National Institute of Mental Health Grant MH-08667: Socio-Behavioral Study Center for Mental Retardation, Pacific State Hospital, Pomona, California; and the National Institute of Child Health and Human Development, HD 02712, Research in Mental Retardation, Mental Retardation Center, University of California at Los Angeles, Los Angeles, California.

so they were selected to serve as the experimental message. Hospital administrators determined the specific instructions to be conveyed, and the writers then prepared the directive in two versions. One, which used the original fire rules as a model, conformed to the wording, layout, and reading level *traditionally* used in the formal messages of the organization. In the other, the message was *simplified* in that sentences were shorter, there were fewer syllables per word, and layout was utilized to improve readability.

The contrasting style is reflected by the following passages:

Traditional version:	5 blasts from the hospital air horn, followed by a brief interval of silence, and continuing until the five blasts have been repeated four times, will indicate a fire in a patient occupied building.
Simplified sentence:	5 blasts on the air horn means fire in a patient area.

To insure that the content was held constant, a panel of hospital administrators reviewed the two versions. Furthermore, a point-by-point analysis showed that both contained the same basic information.

Readability was determined by the Flesch (3) count. Table 1 shows the similarity between the original fire rules and the revised traditional version as contrasted with the revised simplified version. The traditional version was shown by t tests to have significantly more words per sentence ($p < .001$) and significantly more syllables per 100 words ($p < .05$) than the simplified version.

TABLE 1
FLESCH COUNTS OF THREE VERSIONS OF HOSPITAL FIRE RULES

Measure	Original	Revised traditional	Revised simplified
Reading ease score	57.2	51.5	73.4
Number of words	—	532	325
Number of sentences	—	32	49
Average sentence length	17.1	16.6	6.6
Syllables per 100 words	156.3	163.7	149.8
Human interest score	28.8	23.2	32.2
Personal words per 100	.4	.2	3.4
Personal sentences per 100	71.8	71.8	63.3

Reducing the word length was difficult because of the high proportion of monosyllabic terms in both documents. Sixty-four per cent of the words in the traditional version had one syllable. Nineteen per cent had three or more. In the simplified version 72 per cent of the words had one syllable, but only 8 per cent had more than two.

2. *Criteria of Communication Effectiveness*

The selection of criteria for assessing communication effectiveness posed some problems. To be effective, the fire rules would have to be read and understood. Since their ultimate purpose was to prescribe behavior expected of the hospital employees in case of fire, the best criterion of effectiveness would have been a test of performance to see how closely the employees followed the instructions which they contained. This type of evaluation was obviously impractical.

It was decided that communication effectiveness would be assessed by two criteria: (*a*) The frequency with which the two versions of the directive were read. Readership is not an ideal criterion, but it was thought that a more readable, simplified message might attract a higher proportion of readers and thus enhance communication. (*b*) The employees' comprehension, as measured by their ability to answer questions based on the directive. A questionnaire was developed which, in its final form, contained 35 items. The first 11 supplied demographic data and information about the employee's job level and work location, while the remainder were based on the information contained in the directive.

3. *Subjects and Procedure*

The personnel of the Nursing Services Department were divided into two groups of approximately equal size. The traditional version of the fire rules was distributed to one group ($N = 199$) and the simplified version to the other ($N = 190$). Chi square was used to compare the two groups with respect to age, sex, education, length of employment, and job level, and the only significant difference found was with respect to sex. This difference proved to be inconsequential, however, since none of the substantive findings was significantly related to sex.

The procedure was devised in such a way that the employees would have an equal opportunity to read the two versions. The fact that a study was being conducted was concealed in order to minimize distortions that might otherwise have occurred. The new fire rules were issued at noon one day through the hospital mail system, and at mid-afternoon on the following day a group of research assistants simultaneously went to 34 work locations, 33 wards, and a class being conducted for 70 nursing personnel. They administered the questionnaire to the daytime personnel as they were completing their shift and to the afternoon personnel as they arrived for work. No one realized that two versions of the fire rules had been issued until they were told so by the researchers.

C. RESULTS

Chi square was used to test the difference in the proportions of employees who had read the two versions of the fire rules. The traditional version had been read by 35.7 per cent of the subjects who received it, while the corresponding figure for the simplified version was 42.8 per cent. This difference was not significant.

To evaluate comprehension, an analysis of variance was carried out on the questionnaire scores of the two groups of readers and a third group, comprised of those who had read neither version. The mean number of correct responses for the readers of the simplified version was 14.5; for the readers of the traditional version, 13.4; and for the nonreaders, 10.3. The overall difference among the groups was highly significant ($p < .001$).

A series of t tests was applied to evaluate the significance of the differences between the groups taken two at a time. The mean scores for both groups of readers were significantly higher than the mean score for the nonreaders ($p < .001$), and the mean for the readers of the simplified version was significantly higher than the mean for the readers of the traditional version ($p < .05$).

D. DISCUSSION

The present study closely paralleled one conducted by Coleman (2), who studied college students' comprehension of documents that contained sentences averaging 38.7, 23.2, and 15.4 words in length. The two versions of the directive used in the present study averaged only 16.6 and 6.6 words per sentence, but the level of education of the subjects was lower than that of Coleman's subjects. Both studies found that comprehension of shorter sentences was better than that of longer sentences, but that the magnitude of the difference was small. It appears doubtful that reading ease would be greatly improved by shortening the sentences still further.

We may speculate that any organization has an upper limit to the message complexity that can be tolerated, and that this limit is probably higher than most people believe. If the complexity of a message falls below the limit, further simplification produces little change in readership and comprehension.

One of the most interesting, though tangential, findings was that less than half the employees read the experimental message, regardless of the style in which it was prepared. This fact becomes more pointed when it is realized that employees knew that the message was of critical concern to the hospital administration. Apparently written messages must be followed by meetings, conferences, training sessions, and other forms of redundancy to insure that the information they contain reaches the employees who must act upon this information.

E. SUMMARY

This study tested the influence of reading ease on the effectiveness of formal written communications in an organization. Messages were prepared in contrasting styles. One was a traditionally worded interdepartmental directive with a relatively low reading ease score as determined by the Flesch count. The other conveyed the same information, but had a high reading ease score. Word length as measured by the number of syllables was shortened but not as drastically as sentence length.

The research transpired as a part of the normal operations of a large state hospital. The messages were released to 389 employees who were divided into two similar groups. Communication effectiveness was determined by comparing the accuracy with which personnel understood the message. Comprehension of the simplified version was significantly better than that of the traditionally worded directive, but the magnitude of the difference was small.

REFERENCES

1. BELL, G. B., & HOFFMAN, B. J. Attitudes toward the inter-office memo as a communication device. *J. Soc. Psychol.*, 1965, **66**, 15-21.
2. COLEMAN, E. B. Improving comprehensibility by shortening sentences. *J. Appl. Psychol.*, 1962, **46**, 131-134.
3. FLESCH, R. A new readability yardstick. *J. Appl. Psychol.*, 1948, **32**, 221-233.
4. MATHIES, L. H. The Playscript Procedure: A New Tool of Administration. New York: Office Pub., 1961.

The Neuropsychiatric Institute
UCLA Center for the Health Sciences
Los Angeles, California 90024

Published as a separate and in *The Journal of Psychology*, 1968, 68, 55-61.

BRIGHTNESS AND SATURATION CHANGES RESULTANT FROM TEMPORAL MANIPULATION OF SPECTRAL PHOTIC INPUT FOR NORMAL AND COLOR-DEFICIENT OBSERVERS*[1]

Laboratory for the Study of Vision and Related Sensory Processes, Michigan State University

RICHARD J. BALL AND S. HOWARD BARTLEY

A. INTRODUCTION

Brightness changes resultant from manipulation of the temporal aspects of photic input to the visual system have been studied for a number of years by Bartley and colleagues (6, 7, 8, 10). In recent years, hue and saturation changes resulting from stimulus tuning have also been extensively investigated (1, 2, 3, 4, 9, 11). Brightness enhancement is the term applied to the phenomenon of increased brightness from certain intermittency rates (generally in the range of about five-12 cycles per second). Many of these same conditions that produce increased brightness and color effects for achromatic input also produce marked hue and saturation changes for spectral inputs. Total desaturation of a pure spectral band around 500 nm can be produced, as can large hue shifts and desaturation for many wavebands throughout the spectrum. It was found that red-green color deficiency can be introduced by viewing pseudoisochromatic plates under certain forms of intermittent illumination (5, 12). Interest has arisen as to what brightness, hue, and saturation effects would be produced with various types of color deficient observers. This present study reports the brightness and saturation data for one color-normal and three color-deficient observers.

B. APPARATUS

The apparatus has been extensively described previously (1, 3). It basically consists of a prism monochromator that produces a high spectral purity bipartite field. Both halves of the field have the same spectral composition. The photic intensity of one-half of the field can be varied with a rotary crossed polaroid.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 16, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Investigation supported in whole by Public Health Service Research Grant NB 05260-04 from the National Institutes of Neurological Diseases and Blindness.

The other half of the field is made intermittent by the use of an episcotister. Selective and independent manipulation can be achieved for the variables of target wavelength, target size, target intensity level, pulse-to-cycle fraction (PCF) of the intermittent half, rate of intermittency, and photic intensity level of the target-surround field.

C. Procedure

In this particular study a target size of 6°, an intermittency rate of 9.8 cps, an equal luminosity target photic intensity of 432 candles per square foot, and a target-surround field of zero intensity were utilized. The manipulated variables were wavelength composition of the target field and pulse-to-cycle fraction of the intermittent half of the field. Twelve wavebands between 480 nm and 680 nm were used. Pulse-to-cycle fractions between 1/16 and 3/4 were utilized.

Four observers were included in this study. Their color vision was evaluated on the basis of the A. O. H-R-R Pseudoisochromatic Plates. One observer (normal) made no errors on this series and was classified as color normal. A second observer (deutan) was classified as deuteranomalous on the following basis: five errors out of six symbols in the red-green screening series, eight errors out of nine symbols in the deutan diagnostic series, three errors out of nine symbols on the protan diagnostic series, and only one error out of 12 symbols in the total blue-yellow series. The third observer (protan) was classified as protanomalous on the following basis: five errors on the red-green screening series, eight errors on the protan diagnostic series, and no errors on the total blue-yellow series. The fourth observer (mixed) was left unclassified on the following basis: five errors on the red-green screening series, six errors on the protan diagnostic series, five errors on the deutan diagnostic series, and no errors on the total blue-yellow series. While this observer was obviously red-green deficient, a valid decision between protan or deutan could not be made.

For each experimental situation of a particular waveband and PCF combination, the observer made five ascending and five descending brightness matches. These were made by varying the photic intensity of the steady half of the bipartite field with the rotary polaroid arrangement until its brightness matched the brightness of the intermittent half. Thus, each point plotted on Figure 1 is a mean of at least 10 brightness matches. Most of the curves shown are averages of several trials so that each point is a mean of as many as 40 brightness matches.

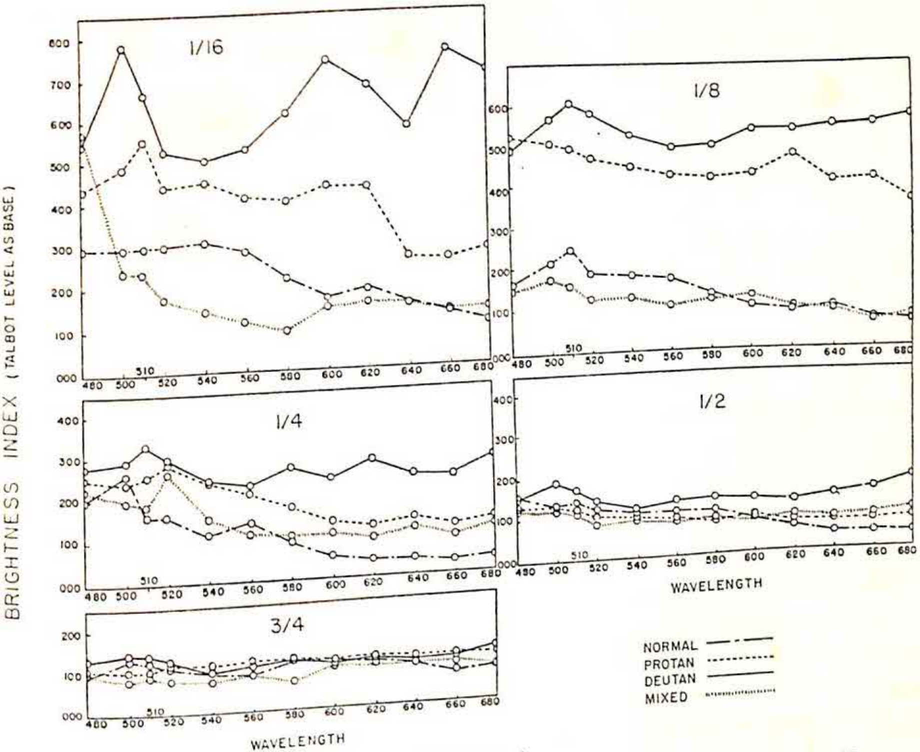


FIGURE 1

Brightness Behavior in Terms of Brightness Index (See Text) for One Normal and Three Color-Deficient Observers, for Five Pulse-to-Cycle Fractions (1/16 to 3/4), and Various Wavelengths

The temporally induced desaturation of the intermittent target-half was judged in relation to the highly saturated steady target-half on the basis of a nine-point subjective scale. This scale ranged from "no desaturation" where the two target-halves were identical in saturation to "total desaturation" where the intermittent target-half appeared totally achromatic. These results are shown on Figure 2.

At all times the observer maintained fixation on a narrow black bar which separated the two halves of the bipartite field. Data have been collected for trials going progressively from bands of shorter wavelength to those of longer wavelength and for trials going progressively from longer to shorter. No great differences in the data collected by these two approaches were found. The data shown in this article are for the shorter to longer wavelength progression.

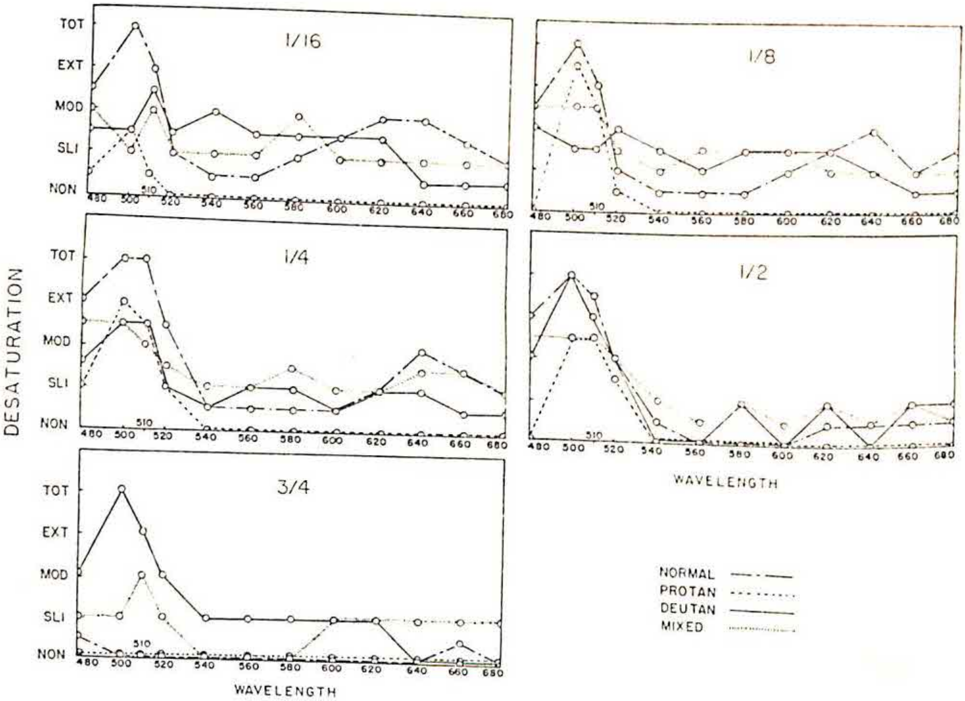


FIGURE 2
AMOUNTS OF DESATURATION PRODUCED BY INTERMITTENT STIMULATION AT 9.8 CPS FOR ONE NORMAL AND THREE COLOR-DEFICIENT OBSERVERS, FOR FIVE PULSE-TO-CYCLE FRACTIONS (1/16 TO 3/4), AND VARIOUS WAVELENGTHS

D. RESULTS

Figure 1 depicts the brightness data for the four observers. The abscissa is wavelength and the ordinate is brightness index. Brightness index has been defined and discussed previously (3), so will be only briefly described here. It is a ratio between obtained and expected brightness with expected brightness being determined on the basis of photic input per cycle in much the same way that Talbot levels are determined. A brightness index of 100 is that point at which obtained brightness equals expected brightness and would be equal to a Talbot expected brightness level for that PCF.

Several points should be made concerning the data of Figure 1.

Color normal: For the color-normal observer (normal), the brightness index tends to decrease as wavelength increases for PCF's up to 1/2 with a peak brightness index in the region of 500 nm. Brightness index for the shorter wavelengths tends to decrease as PCF is increased. For all PCF's the shorter wavelengths yield brightness greater than expected, and for all

PCF's except 1/16 the longer wavelengths yield brightness less than expected in that the brightness index falls below 100. A PCF of 3/4 yields brightness index close to expected brightness for all wavebands.

Deutan: The deutan observer is similar to the color-normal observer in many aspects of his brightness data. There is a slight tendency for brightness index to decrease as wavelength increases, a tendency for brightness peaking around 500 nm, and a PCF of 3/4 yielding brightness of about an expected level for all wavebands. There are also some differences. The deutan tends to have an increase in brightness for the longest wavebands and at no time does the index fall below the 100, or expected level. The deutan has at almost all points a higher brightness index than any of the other three observers. However, the possibility of a judgment criterion difference cannot be totally eliminated with the deutan, who was a relatively inexperienced observer, matching more predominantly to the bright phase of the flicker cycle.

Protan: The protan observer is similar to the color-normal observer with brightness index decreasing as wavelength increases. Brightness index is highest for a PCF of 1/8 with progressive decrease as PCF increases. The 3/4 PCF yields about expected brightness for all wavebands and only with PCF of 1/2 does the index fall below the expected level of 100 with the longer wavelengths. As with the deutan observer, the protan has very high brightness matches for the shorter PCF's which might involve a criterion judgment on bright phase matching.

Mixed: The mixed observer also shows the tendency of the other observers for brightness index to decrease as wavelength increases. PCF's of 1/2 and 3/4 yield essentially expected brightness levels. A PCF of 1/4 has the highest brightness index. The one excessively high point of 480 nm and 1/16 PCF must be viewed with some doubt as a possible criterion shift. Unfortunately this observer was not available to recheck this point.

Figure 2 depicts the desaturation data for the same four observers.

Color normal: For all PCF's except 3/4 there is a strong desaturation peak around 500 nm where total achromaticity is temporally induced in a highly saturated spectral waveband. A second desaturation region of less magnitude occurs around 640 nm and decreases as PCF increases. For PCF of 3/4 there is essentially no desaturation for any waveband. The general trend is for temporally induced desaturation to be great around 500 nm, very slight at 540-560 nm, moderate around 640 nm, and slight again at 680 nm.

Deutan: For the deutan observer the general trend is for desaturation to decrease as wavelength increases. The deutan differs from the color normal in that greatest desaturation occurs with PCF's of 1/2 and 3/4. As with the normal observer, total desaturation occurs only at 500 nm, but there is no second desaturation region around 640 nm.

Protan: For the protan observer there is a very sharp desaturation peak around 500 nm with little desaturation at 480 nm and no desaturation above 520 nm. Maximum desaturation occurs for the intermediate PCF's with little at PCF of 1/16 and none at PCF of 3/4.

Mixed: For the mixed observer the desaturation decreases as wavelength increases and the intermediate PCF's yield the greater desaturation. There is a slight peaking tendency around 500 nm.

E. Discussion

There is a marked similarity in the shape of the brightness curves for all four types of observers. To a lesser extent this similarity holds for the desaturation curves. Very definitely these four observers are much more alike in their brightness and saturation responses than in their responses on a hue test, such as pseudoisochromatic plates. It would appear that the neural components involved in brightness are not necessarily the same as those involved in hue discrimination if color deficiency is the result of abnormal mixtures and proportions of retinal cone elements.

F. Summary

Brightness indexes and desaturation curves for one normal and three color-deficient observers were obtained for five pulse-to-cycle fractions of intermittent stimulation for various wavelengths across the visible spectrum. All observers showed changes in response for the variables used. All observers showed marked similarity in brightness response, but a lesser similarity in the desaturation responses.

References

1. Ball, R. J. An investigation of chromatic brightness enhancement tendencies. *Amer. J. Optom.*, 1964, **41**, 333-361.
2. Ball, R. J., & Bartley, S. H. Effects of temporal manipulation of photic stimulation on perceived brightness, hue, and saturation. *Amer. J. Optom.*, 1965, **42**, 573-581.
3. ———. Changes in brightness index, saturation, and hue produced by luminance-wavelength-temporal interactions. *J. Opt. Soc. Amer.*, 1966, **56**, 695-699.
4. ———. Further investigations of color perception under temporal manipulation of photic stimulation. *Amer. J. Optom.*, 1966, **43**, 419-430.
5. ———. The induction and reduction of color deficiency by manipulation of temporal aspects of photic input. *Amer. J. Optom.*, 1967, **44**, 411-418.
6. Bartley, S. H. Subjective brightness in relation to flash rate and the light-dark ratio. *J. Exper. Psychol.*, 1938, **23**, 313-319.

7. ———. Visual sensation and its dependence on the neurophysiology of the optic pathway. *Biol. Sympos.*, 1942, **7**, 87-106.
8. ———. Brightness enhancement in relation to target intensity. *J. of Psychol.*, 1951, **32**, 57-62.
9. BARTLEY, S. H., & BALL, R. J. The study of target size in production of brightness, hue, and saturation for intermittent photic stimulation in the red end of the visible spectrum. *J. of Psychol.*, 1966, **64**, 241-247.
10. BARTLEY, S. H., PACZEWITZ, G., & VALSI, E. Brightness enhancement and the stimulus cycle. *J. of Psychol.*, 1957, **43**, 187-192.
11. NELSON, T. M., BARTLEY, S. H., & FORD, Z. The brightness of part-spectrum targets stimulating the eye parafoveally at intermittency rates below fusion. *J. of Psychol.*, 1963, **55**, 387-396.
12. NELSON, T. M., BARTLEY, S. H., & MACKAVEY, W. R. Responses to certain pseudo-isochromatic charts viewed in intermittent illuminance. *Percept. Motor Skills*, 1961, **13**, 227-231.

Laboratory for the Study of Vision
and Related Sensory Processes
Department of Psychology
Michigan State University
East Lansing, Michigan 48823

Published as a separate and in *The Journal of Psychology,* 1968, **68**, 63-71.

COHESION IN MARINE RECRUIT PLATOONS*[1]

Navy Medical Neuropsychiatric Research Unit, San Diego, California

LCDR Paul D. Nelson, MSC, USN; and CDR Newell H. Berry, MSC, USN

A. Introduction

The formal structure of a group, as evidenced through well-defined positions and roles, is invariably accompanied by some degree of informal structure reflecting the interpersonal sentiments among group members. Such psychological affinity among group members is often referred to as cohesiveness, a group property characterized by feelings of "we-ness" and *"esprit de corps."*

From a rational point of view, the formal structure of a group tends to be predominantly task-oriented, while the informal aspects of group structure tend more often than not to provide social-emotional ties among group members. While both the task and social-emotional aspects of group life are important, they are not necessarily compatible. Informal structure may, for example, either facilitate or impede the effective functioning of the formal group structure. Lott and Lott (4) point out, in their recent review of the literature, that there is no unequivocal relationship between group cohesiveness and task effectiveness. Nevertheless, it is important to understand the nature and determinants of cohesiveness, particularly in military units within which teamwork and interdependency are so often crucial.

A rather insightful account of the way in which cohesiveness functions is provided by Little (3) in his analysis of the "buddy" system in combat platoons. The "buddy" has two connotations, one of which refers to a close relationship between two individuals and the other to a more diffuse affinity among all members of a group. The former serves the purpose of providing emotional and tactical support for the specific individuals involved, while the latter functions in the form of providing limits for individual group members'

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 19, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Report Number 66-26, supported by the Bureau of Medicine and Surgery, Navy Department, under Research Work Unit MF 022.01.03-9001. The opinions or assertions contained herein are the private ones of the authors and are not to be construed as official or as necessarily reflecting the views of the Department of the Navy. The contributions of Mr. Victor Burton and Michael McNally, HMC, USN are gratefully acknowledged.

behavior relevant to accomplishing the mission of the group. Festinger, Schachter, and Back (1) similarly refer to the value of group cohesiveness in establishing and reinforcing standards of behavior.

In military basic training programs, since group cohesion and *esprit de corps* are partially training goals in their own right, the informal structure of recruit platoons is of more than passing interest. The relative absence of role differentiation among recruits, furthermore, renders group structure primarily a function of informal rather than formal ties. It is within this context that the present study attempts to describe the nature of cohesiveness within a sample of Marine recruit training platoons and to ascertain the personal characteristics and performance patterns of platoon members associated with variations in such cohesion.

B. METHOD

1. *The Sample*

A random sample of 24 recruit training platoons was drawn from a set of 191 such platoons which had formed at the two Marine Corps Recruit Depots over a 12-month period. This sample was obtained in connection with a larger study of adjustment of Marines during first enlistment. The median platoon size in the sample was 68 men at the end of the second week and 60 at the end of the tenth week of an 11-week training program. The average age of all recruits sampled was slightly over 18 years and their average education was approximately 11 years of schooling. The platoons remained relatively intact for the 11-week basic training program. A small proportion of platoon members were either dropped back or discharged during basic training.

2. *Cohesiveness Measure*

A cohesiveness measure was obtained from peer nominations of friendship, each platoon member having nominated the five men in his platoon whom he liked best. Both Lott and Lott (4) and Tagiuri (6) suggest that interpersonal attraction, in the sense of like-dislike, is a central dimension of group cohesion, particularly when role differentiation is not too great.

Although more sophisticated analyses of social structure include the measurement of both direct and indirect interpersonal choices (2), only direct reciprocal choices were included in the present study. This procedure reduced the total number of choices for analysis to the strongest interpersonal ties. Ideally, the most cohesive group would be one in which all group members are linked by mutual positive choices.

Since the present study was based upon a limit of five choices per member, it was not possible for every member to be linked reciprocally to every other member in the sociometric sense. It was, however, possible for great variations to exist in sociometric structure across groups.

The present measure of cohesiveness in each platoon was constructed in such a way as to account for (*a*) the proportion of group members receiving reciprocal nominations, (*b*) the number of reciprocal choices observed in proportion to the total number of such choices possible, and (*c*) the proportion of group members included in the largest clique or substructure (in this study a clique is any reciprocal choice structure involving two or more individuals).

The cohesion index was obtained as follows:

$$\mathrm{COH} = \frac{N - n_i}{N} \times \frac{n_{rc}}{n_{rcm}} \times \frac{n_{lss}}{N}$$

wherein

N refers to total group size in numbers of members,

n_i refers to the number of members involved in no reciprocal choices,

n_{rc} refers to the observed number of reciprocal choices throughout the total group,

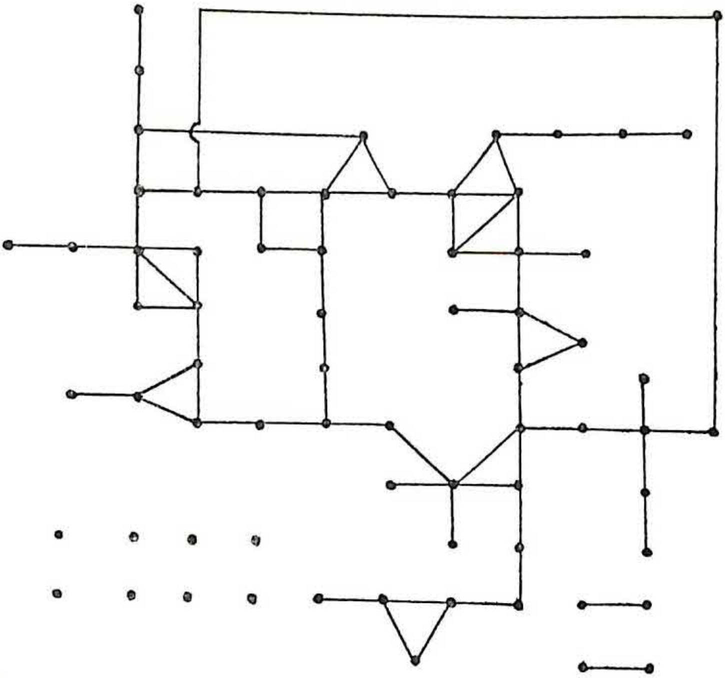
n_{rcm} refers to the maximum number of reciprocal choices possible throughout the total group, and

n_{lss} refers to the number of members in the largest clique or subgroup unconnected by reciprocal choices to other subgroups.

The case representing maximum cohesiveness, in which all group members are interconnected by mutual choices though not necessarily with every other group member, yields a cohesion index (COH) of 1. The case representing minimum cohesiveness, in which no members are reciprocally interconnected, yields a cohesion index value of 0. Each of the components of the cohesion index similarly has a theoretical range of from 0 to 1. Figure 1 illustrates sociograms for a high and a low cohesion platoon. Table 1 provides calculations of cohesion index components for those two platoons.

Peer nominations were obtained during both second and tenth weeks of basic training and an index of platoon cohesion was obtained for each time period. The greatest range of cohesion index values during both time periods was for the proportion of group members in the largest clique. This component was also most highly correlated with the total cohesion index (rho =

High Cohesion Platoon: 69 platoon members



Low Cohesion Platoon: 56 platoon members

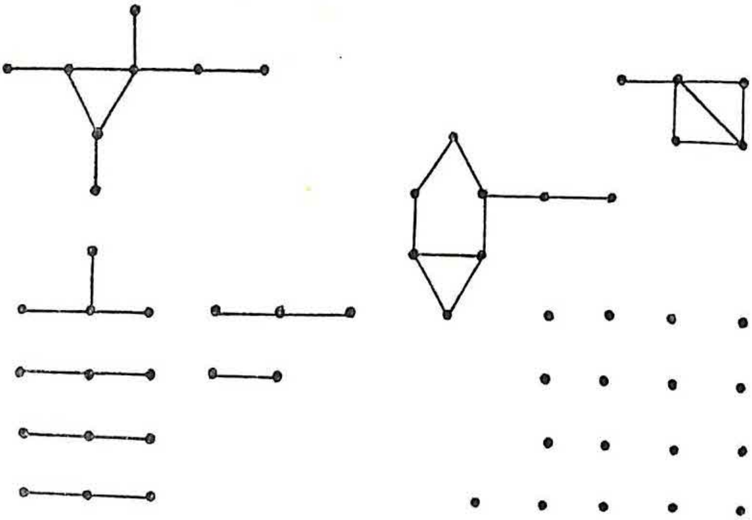


FIGURE 1
SOCIOGRAMS OF MOST AND LEAST COHESIVE PLATOONS
Dots represent platoon members and line represents choice relationships.

TABLE 1

CALCULATION OF COHESION INDEX COMPONENTS FOR THE HIGH AND LOW COHESION PLATOONS OF FIGURE 1

Component	High cohesion platoon	Low cohesion platoon
N	69	56
n_i	8	17
n_{rc}	74	35
n_{rcm}[a]	172.5	140.0
n_{lss}	57	8
Indices		
$\frac{N - n_i}{N}$ (A)	$\frac{61}{69} = .88$	$\frac{39}{56} = .70$
$\frac{n_{rc}}{n_{rcm}}$ (B)	$\frac{74}{172.5} = .43$	$\frac{35}{140} = .25$
$\frac{n_{lss}}{N}$ (C)	$\frac{57}{69} = .83$	$\frac{8}{56} = .14$
COH (A × B × C)	.31	.02

[a] *(Number of choices made per man) (Number of group members)*/2 = $\frac{5(69)}{2}$ and $\frac{5(56)}{2}$.

.95 for both time periods), although all three components of the cohesion index were highly intercorrelated (median value of rho = .74).

C. RESULTS

1. *Change in Cohesion over Time*

Since the unit for study was the platoon and rank ordering among the 24 platoons studied was important for every comparison, the measure of relationship used throughout was rho. Platoon cohesiveness remained relatively stable from second to tenth week of basic training with platoons of highest cohesiveness in second week tending to be most cohesive in tenth week (rho = .35, $p < .05$). In terms of changes in magnitude of the measures of cohesion, the sign-test applied to second- and tenth-week measures revealed no significant differences on overall cohesion, but proportionately fewer platoon members were involved in reciprocal structures during the tenth week than during the second ($p < .05$).

In terms of identity of the specific choices of friendship made, the average platoon member made three (out of five) new choices in tenth week in contrast with choices made during the second week. For both time periods, approximately 40 per cent of the choices were reciprocated. While there were many platoons in which the largest clique contained at least half the members of the platoon, most cliques tended to be relatively small during both time periods. Approximately 60 per cent of all cliques identified by the two-week sociometric consisted of two members, while only about 55 per cent of the cliques at the end of 10 weeks were of that size; at two weeks approximately 28 per cent of the cliques consisted of three to 10 members, while at 10 weeks cliques of that size comprised 30 per cent of the total; and at the two-week period, 10 per cent of the cliques were made up of 11 or more members, but at 10 weeks about 12 per cent of the cliques were of that size. More than half of all cliques in both periods were dyads, suggesting the existence of a two-man buddy system as discussed by Little (3).

2. *Platoon Composition and Cohesion*

One of the possible determinants of group cohesiveness is the extent to which group members share a common set of values and orientations in the situation which brings them together. In the present instance all members were together for the purpose of being trained to be Marines. But the recruits forming these training platoons had come from rather diverse backgrounds with correspondingly different attributes. Data were available on age, education level, type of high school curriculum, size of high school, participation in extracurricular activities, and geographical region of home residence. Since homogeneity of background was of central interest in the present study, estimates of variance were obtained on those attributes and background variables.

Again by rank-order analyses, platoon cohesion for second and tenth weeks separately was studied in relation to variance on personal characteristics of members in the platoons. Age variance was negatively correlated with platoon cohesion during both time periods (rho values of —.43, $p < .05$, and —.55, $p < .01$ for second and tenth weeks, respectively). Variance on education level was also negatively correlated with cohesion but only in the tenth week (rho = —.66, $p < .01$). The only other result of statistical significance was that between geographical region of residence and cohesion. For that analysis, platoons were ranked on the basis of the largest percentage of their members from a common geographical region—recruiting districts serving as regions in this instance. A rho of .37 ($p < .05$) was obtained between that index and cohesion, again in tenth week only. Thus, the more alike platoon members were on age, the more cohesive they tended to be from the start of training. By the end of training, cohesiveness was also related to platoon members being similar on education and coming from similar geographical regions of the country. The latter two attributes are perhaps more related to basic values which in turn are not exchanged on an interpersonal level until a platoon has settled as a group. Age, on the other hand, may be more overtly and immediately reflected in general level of maturity, and thus somewhat more discernable from the earliest stages of a group's life. Similarity in age truly renders platoon members peers, an important aspect of initial adjustment to military training.

No significant relationship was obtained between cohesion and platoon size during either time period, even though there was as much as a 30-man difference between the largest and smallest platoons. By dividing second-week platoon size into tenth-week platoon size, a platoon membership stability ratio was derived which accounted for loss of members from the platoon. That index had a significant positive correlation (rho $= .41$, $p < .05$) with tenth-week, but not second-week, cohesion. Thus early cohesiveness did not seem to result in greater platoon membership stability, but the members of more stable platoons appeared more cohesive by the end of training.

3. *Platoon Performance and Cohesion*

The final set of analyses evaluated cohesion in relation to performance and attitudes of platoon members. There was no complete set of performance data available to reflect achievement of platoons as groups. There were, however, several measures of individual performance which were summarized for each platoon to serve as approximations of group achievement. One of those measures, obtained both early and late in basic training, was the physical proficiency test on which recruits received performance scores. Another test was that of the final rifle qualifications exercise, on which individual members received proficiency scores. Neither second- nor tenth-week cohesion was significantly correlated with either of those two performance measures. Furthermore, on the physical proficiency test, improvement in proficiency from second to tenth weeks was unrelated to cohesion at either time period. In terms of performance subsequent to basic training, such as satisfactory completion of the first two years of service, platoon cohesiveness was again unrelated to performance.

A somewhat different form of behavior studied in addition was that pertaining to platoon member judgments about the future effectiveness of their peers. During both second and tenth weeks, recruits had nominated peers in their platoon who seemed most and least likely to be effective Marines. By correlating those peer nomination status scores with performance two years later (i.e., whether the Marine was still on duty and two-year pay-grade), validity coefficients were obtained for each platoon. Those in turn were ranked and analyzed in relation to platoon cohesion in an attempt to ascertain whether cohesiveness was in any way related to validity of peer group judgments. No significant correlations were obtained.

Finally, analysis was made of the relationship between platoon cohesion and the attitudes of platoon members toward the Marine Corps. During the tenth week of training, attitude data had been obtained from members of three-fourths of the platoons. One item of interest pertained to re-enlistment intention. Other items were pooled to describe attitudes about the Marine Corps in general and also about the calibre of leadership therein (5). No significant correlations were obtained between attitude data and tenth-week cohesion, although platoons below the median on cohesion for both second and tenth week were significantly less favorable on attitudes toward the Corps than other platoons ($p < .05$).

D. Discussion

Since one of the goals of military basic training is to instill the spirit of teamwork and interpersonal dependency in new recruits, it seems of value to know something about the way in which cohesiveness in training platoons develops. With a random sample of 24 Marine recruit platoons, a measure of platoon cohesion was developed for the second and tenth week of training. That measure, based upon interpersonal friendship choices, included constructs of (*a*) the proportion of group members receiving reciprocal nominations, (*b*) the proportion of reciprocal choices observed in relation to the maximum number possible, and (*c*) the proportion of group members in the largest clique.

Platoons remained moderately stable in level of cohesiveness from the second to tenth week of training, even though many individual friendship choices changed and somewhat fewer group members were involved in strong reciprocal interpersonal ties by the end of training. Of all substructures or cliques formed, dyads were proportionately greatest in number. The emotional and tactical support offered in a two-man buddy system may be more rewarding in a recruit-training situation where behavioral limits are already highly structured by the authorities.

Cohesiveness during both time periods was positively related to homogeneity of platoon members' age, and by tenth-week cohesiveness was also positively related to homogeneity of platoon members' education and geographical region of residence. The most cohesive platoons at the end of training tended also to be those platoons whose membership remained most intact through training.

Consistent with other studies of cohesiveness and its correlates, the present results provided no significant relationship between group cohesion and performance measures obtained either during training or during the two years subsequent to training. Such results are perhaps a function of the dimension of interpersonal attraction used to define cohesiveness and also of the type of performance criterion used. In the present study, there was not a criterion measure available which can clearly be said to be a measure of group performance: i.e., performance resulting from the coordinated efforts of all group members. Instead, individual performance measures were used. Cohesiveness, was, nevertheless, moderately related to the attitudes held by recruits about the Marine Corps; those platoons least cohesive throughout training were least favorable in attitude.

E. SUMMARY

Cohesiveness of Marine basic training platoons was observed to be relatively stable over a two-month period, was related to platoon personnel composition and attitudes, but was uncorrelated with performance measures.

REFERENCES

1. FESTINGER, L., SCHACHTER, S., & BACK, K. The operation of group standards. In D. Cartwright & A. Zander (Eds.), *Group Dynamics*. Evanston, Ill.: Row, Peterson, 1953. Pp. 204-222.
2. FORSYTHE, E., & KATZ, L. A matrix approach to the analysis of sociometric data. *Sociometry*, 1946, **9**(4), 340-349.
3. LITTLE, R. W. Buddy relations and combat performance. In M. Janowitz (Ed.), *The New Military*. New York: Russell Sage Foundation, 1964. Pp. 195-224.
4. LOTT, A. J., & LOTT, B. E. Group cohesiveness as interpersonal attraction: A review of relationships with antecedent and consequent variables. *Psychol. Bull.*, 1965, **64**(4), 259-309.
5. NELSON, P. D., & BERRY, N. H. Attitudes of Marines during first enlistment. Report No. 66-21, U.S. Navy Medical Neuropsychiatric Research Unit, San Diego, California, May, 1966.
6. TAGIURI, R. Social preference and its perception. In R. Tagiuri & L. Petrullo (Eds.), *Person Perception and Interpersonal Behavior*. Stanford, Calif.: Stanford Univ. Press, 1958. Pp. 316-336.

Bureau of Medicine and Surgery
Navy Department
Washington, D.C. 20390

Published as a separate and in *The Journal of Psychology,* 1968, **68**, 73-82.

THE COMPARABILITY OF WAIS AND WISC SUBTEST SCORES AND *IQ* ESTIMATES* [1]

Department of Psychology, Marquette University[2]

MOHAMMED Y. QUERESHI

A. INTRODUCTION

A large body of research data concerning the relationships among the Wechsler scales—the Wechsler-Bellevue Scale (WB), the Wechsler Intelligence Scale for Children (WISC), and the Wechsler Adult Intelligence Scale (WAIS)—has accumulated over the past two decades [see the periodic reviews by Guertin *et al.* (2, 3, 4) and Littell (5)]. Most of these studies attempted to verify the comparability of two of the aforementioned Wechsler scales administered either concurrently (a few days to a few months apart) or after substantial delays (ranging from one to six years or more) to normal or exceptional individuals. Although the findings are by no means unanimous, they generally tend to support the following conclusions:

1. The WISC and WBI, for the overlapping ages, yield (*a*) correlations in the .70s and .80s for the FS*IQ;* and (*b*) significantly discrepant V*IQ,* P*IQ,* and FS*IQ,* with V*IQ* on the WISC being higher and P*IQ* and FS*IQ* on the WISC being lower than the corresponding indices on the WBI.

2. The WBI and WAIS yield (*a*) correlations in the .70s and .80s for V*IQ* and FS*IQ* but in the .40s for P*IQ;* (*b*) similar V*IQ;* and (*c*) significantly discrepant P*IQ* and FS*IQ,* with both these estimates on the WBI being larger than their counterparts on the WAIS.

The WISC and WAIS, however, have not yet been compared with each other except for mental retardates or other similarly exceptional groups. Such a comparison, especially if it is based on concurrent administrations, may be questioned on the grounds that the tests are designed for nonoverlapping ages

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 23, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study is based on the data collected as part of a project supported by Grant No. 11115 from the National Institute of Mental Health during 1965-68.

[2] The following abbreviations have been used to designate various WISC or WAIS subtests and *IQ*s throughout the body of this paper: I (Information), C (Comprehension), A (Arithmetic), S (Similarities), V (Vocabulary), D (Digits), PC (Picture Completion), PA (Picture Arrangement), BD (Block Design), OA (Object Assembly), Cod/DS (Coding or Digit Symbol), V*IQ* (Verbal *IQ*), P*IQ* (Performance *IQ*), and FS*IQ* (Full Scale *IQ*).

and, hence, need not be subjected to such investigation. On the other hand, one need not take too seriously, except for normative purposes, the edict that, if a test has not been standardized for a certain age group by its author or publisher, it provides no useful information if such conventions are contravened. If two psychological instruments are devised so that the highest age level encompassed by one borders on the lowest age segment covered by the other, it may be quite revealing to compare and contrast the results obtained by administering both of them either to individuals just below the chronological dividing line or to those just above it. Furthermore, an appropriately designed study, at the same time, can throw some light on the practice and order effects that inhere in such administrations. The present study was intended to attain the foregoing purposes by means of a counterbalanced administration of both the WAIS and WISC to a group of 15 year olds over an average intertest interval of about three months.

B. METHOD

1. *Subjects*

After securing the cooperation of one of the two largest school systems in metropolitan Milwaukee, one of its 12 high schools was selected at random. This school system had an enrollment of about 50,000, from kindergarten through 12th grade, during 1964-65, while the selected high school had a student body of about 2700 during the same period. A total of 124 *S*s, 62 males and 62 females, were randomly selected from all those who happened to be between 14 years, 6 months, and 15 years, 6 months (both ends inclusive) at the time of the first testing with either the WISC or WAIS. The mean age of the total group, on first testing, was 180.1 months, with a standard deviation of 3.56, while means and standard deviations for males and females were 181.9 and 182.2 and 3.58 and 3.54 months, respectively. Chronologically, therefore, the males and females were almost identical ($t = .47$).

No attempt was made to control the nationality of descent or the racial background of the children. The sample consisted of 123 whites and one nonwhite. None of these *S*s suffered from any behavioral or physical handicap.

2. *Procedure*

The selected *S*s were tested during the school hours in a room set aside for this purpose in the school. Sixty-two *S*s (31 males and 31 females) were administered WAIS first and the other 62 (31 males and 31 females) were given WISC initially. All the testing was conducted by two male graduate assistants both of whom were native-born, white Americans between 21 and 25 years in age.[3] Each examiner tested and retested the same 62 *S*s with the two instruments, but the initial assignment of *S*s was purely incidental. With the exception of one subtest, Mazes, both tests were administered *in toto.* The mean test-retest interval was 100 days, with a standard deviation of 13.6.

In the case of seven subtests (I, C, A, S, V, D, and PC), testing was carried out to the point of six consecutive failures instead of the discontinuance limits set down by Wechsler (9, 10). However, the data for the present study were obtained in complete accord with the discontinuance procedures laid down by Wechsler and are comparable to scores that might have been obtained had these seven subtests been discontinued exactly at the points recommended by Wechsler. Other than the foregoing, there was no modification of the cutoff rules regarding the other four subtests nor of any other rule or dictum pertaining to the administration or scoring of any of the 11 subtests either in WAIS or WISC.

3. *Treatment of Data*

For a number of reasons which are specified later, the investigator decided to carry out some of the analyses on both the raw and scaled scores and others on either the raw or scaled scores. Thus, the *IQ* data as well as the subtest scores were subjected to the following treatment: (*a*) With the employment of an analysis of variance procedure recommended by Grant (1), 22 analyses (11 based on raw and 11 on the scaled scores) for the subtests and three analyses (one each for V*IQ*, P*IQ*, and FS*IQ*) for the *IQ* estimates were completed so that the main effects due to order of presentation, practice, and test material could be evaluated unambiguously. (*b*) Utilizing scaled scores only, Pearson *r*s were computed among all 28 variables (11 subtest and three *IQ* scores for each of the two batteries). These correlations, in conjunction with the corresponding means and variances, provided both the necessary and sufficient basis for evaluating the comparability of various subtest scores and *IQ* estimates. Correlations between corresponding subtests (having similar names or content) and *IQ* scores on the WISC and WAIS were corrected for the restriction in the range of talent in the respective distributions by placing the standard deviations equal to 3.0 or 15.0, depending on whether the variables represented subtest scores or *IQ* estimates.

---

[3] Acknowledgment is due to Richard Kindel and Jeffrey Miller for their conscientious and diligent efforts in carrying out the individual testing for this study.

C. Results and Discussion

1. *Differences Between Subtest Scores and* IQ *Estimates*

It is well known that the comparability of tests depends on the equality of their (*a*) means and variances, (*b*) intercorrelations (reliability coefficients) when there are more than two tests involved, and (*c*) correlations (validity coefficients) with one or more criterion measures. Since the scaled scores for the various subtests in the WISC are based on one age group and those in the WAIS on another, it is essential that the equivalence with respect to (*a*) above be judged on the basis of raw scores, whereas the measures on criteria (*b*) and (*c*) may be based either on the raw or scaled scores. Table 1 presents means, standard deviations, and significance values (t-1 and t-2) for the corresponding mean and variance differences based on raw scores on the 11 subtests as well as for the three *IQ* estimates. The means and standard deviations of scaled scores on the same variables also appear in Table 1, since they provide a ready comparison of the performance of the present group with the original standardization data of the WISC and WAIS. Both the WISC and WAIS data in Table 1 indicate that the present sample definitely deviates from the normal population both in intellectual level (higher mean *IQ*s) and variability (lower standard deviations). Taking into consideration criterion (*a*), all subtests except D turn out to be nonequivalent. Although two other subtests, C and PC, have the same means, their variances are significantly different from each other. Among the *IQ*s, P*IQ* and FS*IQ* are not comparable at all, while V*IQ* provides some semblance of equivalence if only the means are considered. One may conclude that, in general, WISC and WAIS subtests and *IQ*s are not comparable when such judgments are based on the equality of means and variances. Furthermore, this conclusion remains essentially unchanged even if the analysis is based on scaled scores (see Table A).[4] Comments concerning criteria (*b*) and (*c*) of equivalence are postponed until after the examination of order and practice effects.

The results of the analysis of variance for the three *IQ* estimates are incorporated in Table 2. It is apparent that the practice effects are highly significant ($p < .01$) for all three types of *IQ*. The order effects, although significant for V*IQ* and FS*IQ*, cannot be interpreted unambiguously unless

---

[4] A five-page document containing Tables A through E has been deposited as Document number 9687 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D.C. 20540. A copy may be secured by citing the Document number and by remitting $1.25 for photoprints, or $1.25 for 35-mm microfilm. Advance payment is required. Make checks or money orders payable to Chief, Photoduplication Service, Library of Congress.

*S*s-within-order effect is discounted by computing new *F*-ratios using order mean squares and *S*s-within-order mean squares as numerators and denominators, respectively. These *F*-ratios of 1.53, .21, and .99 ($dfs = 1, 122$) for V*IQ*, P*IQ*, and FS*IQ*, respectively, are not significant, indicating that there are no order effects independent of significant individual differences within a sequence.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND SIGNIFICANCE VALUES OF THE MEAN (*t*-1) AND VARIANCE (*t*-2) DIFFERENCES FOR SUBTEST SCORES AND *IQ* ESTIMATES

	Means		Standard deviations		*t*s ($dfs = 122$)	
Variable	WISC	WAIS	WISC	WAIS	*t*-1	*t*-2
I	21.1 (11.7)	17.5 (11.0)	2.71 (2.55)	3.75 (1.90)	15.10***	—4.82***
C	17.6 (10.8)	17.5 (10.5)	2.79 (2.74)	3.92 (2.86)	.53	—4.35***
A	12.5 (10.6)	10.5 (9.7)	1.76 (2.70)	2.53 (2.34)	9.94***	—4.54***
S	17.7 (13.0)	15.9 (11.3)	3.36 (2.93)	3.18 (1.73)	5.50***	.65
V	51.8 (11.6)	43.4 (10.1)	7.26 (2.34)	11.51 (1.85)	12.61***	—8.37***
D	11.4 (10.6)	11.1 (10.0)	2.18 (3.01)	2.09 (2.71)	1.69	.62
PC	14.6 (11.3)	14.5 (10.5)	2.35 (2.80)	2.87 (2.03)	.40	—2.45*
PA	36.2 (10.7)	23.5 (10.0)	6.93 (2.76)	5.23 (2.23)	17.48***	3.16**
BD	37.2 (11.0)	34.2 (10.5)	13.18 (3.63)	7.90 (2.57)	3.40***	7.00***
OA	26.6 (11.3)	32.5 (10.7)	4.38 (3.16)	6.63 (3.00)	—12.25***	—5.73***
Cod/DS	68.6 (13.8)	57.4 (10.8)	10.34 (2.90)	10.54 (2.43)	18.27***	.28
V*IQ*	108.8	107.6	10.71	8.65	1.94	3.66***
P*IQ*	111.2	104.9	13.47	10.38	7.65***	3.97***
FS*IQ*	110.9	107.0	10.86	8.26	7.18***	4.91***

Note: Numbers in parentheses represent estimates based on scaled scores; all other computations for subtests are based on raw scores. Each *t*-1 is the square root of the corresponding *F*-ratio for the main effect due to subtests or *IQ*s in the analysis of variance described in the text. The *t*-2 values have been computed with due regard for the correlation between variances [see McNemar (6, p. 246)]. A negative *t* indicates that the WAIS estimate on that variable is larger than the corresponding WISC score.

* $p \leqslant .05$.
** $p \leqslant .01$.
*** $p \leqslant .001$.

TABLE 2

ANALYSIS OF VARIANCE OF THE V*IQ*, P*IQ*, AND FS*IQ* OF THE WISC AND WAIS

		Mean square			*F*-ratio		
Source	*df*	V*IQ*	P*IQ*	FS*IQ*	V*IQ*	P*IQ*	FS*IQ*
Tests	1	85.97	2396.94	944.56	3.76	58.49***	51.62***
Practice	1	184.69	1028.31	570.06	8.08**	25.10***	31.15***
Order	1	252.03	49.69	161.28	11.02**	1.21	8.81**
*S*s-within-order	122	164.62	241.63	163.41	7.20***	5.90***	8.93***
Error	122	22.87	40.98	18.30			
Total	247						

** $p \leqslant .01$.
*** $p \leqslant .001$.

An examination of the effects of practice and order on scores on each of the 11 subtests separately would permit the identification of test material susceptible to either or both of these factors. The data in Table 3 specifically provide such information. It is apparent that C, V, and D of the verbal and PC and PA of the performance section are completely insusceptible to any practice effects. Two of the remaining verbal subtests (I and S) exhibit only a moderate influence of practice ($.03 < p < .05$), while all three remaining performance subtests (BD, OA, and Cod/DS) show definitely greater amenability to the effect of repeated adminstrations over a three-month period. This fact is responsible for the larger *F*-ratio for the practice effects on P*IQ* than that for V*IQ*, and is consistent with previous findings [see Guertin *et al.* (4), and Quereshi (8)]. The order effects, when *S*s-within-order differences are held constant, reach the minimum acceptable level of significance ($p < .05$) only for C, A, PC, and OA subtests (*F*-ratios being 5.99, 5.67, 6.65, and 4.57, respectively, for 1, 122 *df*s). For each of these four subtests, somehow, it makes a difference whether its counterpart in the WISC or WAIS is administered first. Corresponding analyses utilizing scaled scores lead to the same overall conclusions, though there are some variations with respect to two or three subtests (see Table B).[5]

2. *Intercorrelation of Subtests and* IQ*s*

So far only criterion (*a*) of comparability has been applied to the WISC and WAIS. Criterion (*b*) ideally requires more than two comparable forms, but since only two forms of each of the 11 subtests or their composites are available in the present study—and this is feasible only if one regards, on a hypothetical basis, subtests like I, C, A, etc., of the WISC as parallel forms of the corresponding subtests I, C, A, etc., of the WAIS—their equivalence has to be judged on the basis of a single coefficient (delayed parallel-form reliability) for each of the subtests and *IQ* scores. Table 4 presents the actual correlations (r_{12}) between the corresponding WISC and WAIS subtests and *IQ*s as well as those indices (R_{12}) that have been obtained after making appropriate adjustment for the restriction in the range of talent in both variates. The adjusted coefficients (R_{12}) clearly indicate that the V*IQ* and FS*IQ* for the two batteries are substantially the same, since their R_{12} indices are as high as their internal consistency estimates based on the original standardization data [see Wechsler (9, p. 13) and Wechsler (10, p. 13)]. The P*IQ*, despite its substantial R_{12} (.84), fails to satisfy the requirement of correlational equivalence: i.e., having R_{12} as high as the internal consistency indices reported by Wechsler. Individually, only three subtests, I, V, and D, meet the necessary requirement, and the remaining subtests with the exception of PA, in spite of their substantial intercorrelations ($p < .001$), fall short of being correlationally equivalent.

---

[5] See footnote 4.

TABLE 3

ANALYSIS OF VARIANCE OF THE RAW SCORES ON THE 11 SUBTESTS OF THE WISC AND WAIS

Source	*df*	I	C	A	S	V	D
				Mean square			
Subtest	1	787.8	1.8	242.2	189.9	4411.8	4.4
Practice	1	17.6	.2	47.9	30.5	58.1	.2
Order	1	35.6	96.9	36.4	9.7	174.5	.7
*S*s-with-order	122	17.7	16.2	6.4	15.0	157.0	7.6
Error	122	3.5	6.3	2.5	6.3	27.8	1.5
				F-ratio			
Subtest	1	227.91***	.28	98.90***	30.28***	159.00***	2.85
Practice	1	5.08*	.03	19.58***	4.87*	2.09	.13
Order	1	10.31**	15.28***	14.87***	1.54	6.29*	.44
*S*s-with-order	122	5.12***	2.55***	2.62***	2.39***	5.66***	4.94***

Source	*df*	PC	PA	BD	OA	Cod/DS
				Mean square		
Subtest	1	.7	10117.2	537.2	2166.5	7746.0
Practice	1	1.8	.4	1515.2	150.2	1735.2
Order	1	61.0	5.8	82.5	211.5	68.1
*S*s-with-order	122	9.2	42.8	178.4	46.3	181.8
Error	122	4.2	33.1	46.6	14.4	23.2
				F-ratio		
Subtest	1	.16	305.64***	11.54***	150.45***	333.83***
Practice	1	.42	.01	32.54***	10.43**	74.78***
Order	1	14.56***	.18	1.77	14.68***	2.94
*S*s-with-order	122	2.19***	1.29	3.83***	3.22***	7.83***

Note: The *F*-ratios above have been computed using the mean squares before rounding and therefore may deviate, within rounding limits, from the *F*-ratios based on the mean squares reported above.

* $p \leqslant .05$.
** $p \leqslant .01$.
*** $p \leqslant .001$.

A comparison of r_{12} indices for the three *IQ*s in Table 4, with the corresponding coefficients representing the relationship between WAIS and WBI in another study (7), leads to the conclusion that WAIS and WISC are more alike than WAIS and WBI. Furthermore, if r_{12} in Table 4 and the corresponding coefficients from Neuringer's study (7, p. 760) are converted into their appropriate significance values (see Table C),[6] it becomes apparent that most of the WAIS subtests correlate more significantly with the corresponding WISC variables than they do with the respective WBI measures. Thus, on the basis of intercorrelations, WAIS is more comparable to WISC than to WBI, if such a statement has to be made at all.

It may be recalled that previously three separate criteria of equivalence were enumerated, two of which have been applied to the available data. The third criterion (i.e., the equality of validity coefficients) is applicable only if for each subtest in the WAIS or WISC one or more behavioral measures (other than the subtest or its counterpart) are available. The purpose here, however, is not as much to predict an external criterion as it is to demonstrate that two (or more) comparable subtests are uniformly correlated with one or more measures which are psychologically different from either one of them. Thus, one may use, for any one WISC subtest, the mean r based on its correlations with all the WISC subtests except itself and the mean r for the corresponding WAIS subtest based on its correlations with all the WISC subtests except its counterpart as a basis for judging the uniformity of

TABLE 4

DELAYED PARALLEL-FORM RELIABILITY (r_{12}), MEAN CORRELATIONS ($\bar{r}$-1 AND $\bar{r}$-2), AND ts FOR THE DIFFERENCES BETWEEN MEAN CORRELATIONS FOR THE WISC AND WAIS VARIABLES

Variable	r_{12}	R_{12}	$\bar{r}$-1	$\bar{r}$-2	t
I	.67	.86	.24	.27	0.35
C	.48	.54	(.13)	.23	1.15
A	.44	.57	.20	.21	.12
S	.34	.54	.27	.24	.33
V	.76	.93	.28	.30	.24
D	.67	.67	(.09)	(.12)	.34
PC	.43	.61	.23	.27	.47
PA	.12	.17	(.12)	.23	1.27
BD	.54	.53	.23	.18	.58
OA	.57	.55	.19	.18	.12
Cod/DS	.67	.76	(.11)	(.09)	.23
VIQ	.76	.94			
PIQ	.68	.84			
FS*IQ*	.78	.95			

Note: R_{12} represents r_{12} corrected for the restriction in the range of talent in both variates. All r_{12}, with the exception of that for PA, are significant beyond .001. All $\bar{r}$-1 and $\bar{r}$-2 indices, except those in parentheses, are significant at or beyond .05. The sign of the ts has been ignored, and they have been computed with due regard to the intercorrelation of mean rs. All of the above indices are based on 124 *S*s.

---

[6] See footnote 4.

"validity" coefficients.[7] Column 4 in Table 4 represents the average correlation ($\bar{r}$-1) of each WISC subtest with all other WISC subtests except itself, while column 5 in Table 4 provides the average correlation ($\bar{r}$-2) of each WAIS subtest with all other WISC subtests except its counterpart. No such estimates are reported for the WISC or WAIS *IQ*s, since the foregoing rationale does not apply to composite scores for which no external variables are available. Complete data on correlations of all the WISC or WAIS subtests and *IQ*s are presented in Tables D and E.[8]

A comparison can now be made between $\bar{r}$-1 and $\bar{r}$-2 estimates for any one of the 14 variables in Table 4. The *t*-values approximating the significance of the differences between the corresponding pairs of mean correlations, for the 11 subtests, are given in the last column of Table 4. Since all *t*s are insignificant (the largest t of 1.27 has $p > .20$), all 11 subtest pairs can be adjudged to have uniform "validities." However, in order to be considered equivalent, any two (or more) instruments have to satisfy all three criteria of comparability. In the present study, D is the only subtest which meets all of the criteria.

A careful examination of the past literature (through 1966) in this area indicates that not a single study attempted to apply all of the statistical criteria of equivalence to the comparison of any of the Wechsler scales. Although a majority of these studies, as Guertin *et al.* (3, p. 388) point out, suffer from several methodological inadequacies, even the few that employed adequate experimental design [e.g., Neuringer (7)] failed to apply the appropriate statistical criteria of comparability. Viewed in this light, the present study, besides its specific findings regarding WISC and WAIS, seems to have illustrated clearly the appropriate design and analysis that need to be employed by studies of this type.

D. SUMMARY

The present study investigated the comparability of the WISC and WAIS subtest scores and *IQ* estimates for a group of 124 (62 males and 62 females) 15 year olds. The scales were administered in a counterbalanced fashion with an interval of about three months between the testings.

Results indicated definite practice effects for the *IQ*s as well as for a number of individual subtests. When the three criteria of equivalence were

---

[7] The WISC subtests were used as the "validating" variables, since their scaled scores, because of the age level of the present sample, are more appropriately comparable to the Wechsler standardization data than the WAIS scaled scores.

[8] See footnote 4.

applied, it was found that (*a*) most subtests and *IQ* scores failed to meet the first criterion (i.e., equality of means and variances), (*b*) two *IQ*s (V*IQ* and FS*IQ*) and three subtests (I, V, and D) adequately met the second criterion (high delayed parallel-form reliability), and (*c*) all subtests satisfied the third criterion (uniformity of "validity" indices). On the basis of the correlational criteria, (*b*) and (*c*), WAIS turned out to be more akin to WISC than to WBI.

REFERENCES

1. GRANT, D. The statistical analysis of a frequent experimental design. *Amer. J. Psychol.*, 1949, **62**, 119-122.
2. GUERTIN, W. H., FRANK, G. H., & RABIN, A. I. Research with the WB Intelligence Scale. *Psychol. Bull.*, 1956, **53**, 235-257.
3. GUERTIN, W. H., LADD, C. E., FRANK, G. H., RABIN, A. I., & HEISTER, D. S. Research with the Wechsler Intelligence Scales for Adults: 1960-65. *Psychol. Bull.*, 1966, **66**, 385-409.
4. GUERTIN, W. H., RABIN, A. I., FRANK, G. H., & LADD, C. E. Research with the Wechsler Intelligence Scales for Adults: 1955-60. *Psychol. Bull.*, 1962, **59**, 1-26.
5. LITTELL, W. M. The Wechsler Intelligence Scale for Children: Review of a decade of research. *Psychol. Bull.*, 1960, **57**, 132-156.
6. MCNEMAR, Q. Psychological Statistics (3rd ed.). New York: Wiley, 1962.
7. NEURINGER, C. The form equivalence between the Wechsler-Bellevue Intelligence Scale, Form I, and the Wechsler Adult Intelligence Scale. *Educ. & Psychol. Meas.*, 1963, **23**, 755-763.
8. QUERESHI, M. Y. Practice effects on the WISC subtest scores and *IQ* estimates. *J. Clin. Psychol.*, in press.
9. WECHSLER, D. Wechsler Intelligence Scale for Children. New York: Psychol. Corp., 1949.
10. ———. Wechsler Adult Intelligence Scale. New York: Psychol. Corp., 1955.

Department of Psychology
Marquette University
617 North Thirteenth Street
Milwaukee, Wisconsin 53233

Published as a separate and in *The Journal of Psychology,* 1968, **68**, 83-86.

PREDICTING THE FULL SCALE WAIS *IQ* OF ARMY BASIC TRAINEES[*]

Morristown, New Jersey

JOSEPH DELUCA

A. INTRODUCTION

During the author's two years of experience in the Mental Hygiene Consultation Division (MHCD) of a United States Army Hospital (USAH),[1] one observation was repeatedly reinforced. This was the fact that psychological evaluations which included a valid estimate of intellectual functioning and capacity were consistently asked for by the other mental health professionals in increasing numbers, while the psychology service staff in most MHCDs was decreasing or remaining the same. The importance of having this information could not be denied in order to make a competent disposition of many cases referred to the MHCD, and there appeared to be a need to make more efficient the psychodiagnostic process relative to intelligence. The purpose of this study, therefore, was to find a way of measuring, in an efficient manner, not only the overall general intellectual functioning of a trainee but also differential cognitive functions, such as verbal comprehension, memory, and perceptual organization.

As a result of a factor analysis on several different populations, Cohen (1) found that the WAIS measures three major factors: verbal comprehension (VC), nonverbal perceptual organization (NVPO), and memory (M). A second-order factor was interpreted as general intelligence (GI). The factor scores are derived by averaging age weighted scores as follows—VC: Information (I), Comprehension (C), Similarities (S), and Vocabulary (V); NVPO: Block Design (BD) and Object Assembly (OA); M: Arithmetic (A) and Digit Span (DS); and GI: all of the subtests above.

By taking the two subtests which correlate most highly with VC (I, V) with the subtests which correlate highest with the other factors (BD, OA for NVPO and A, DS for M), we can get a good test of what the WAIS really measures while saving approximately one-half the total time to admin-

---

[*] Received in the Editorial Office, Provincetown, Massachusetts, on October 24, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Fort Jackson, South Carolina.

ister and score the full WAIS. It was felt that the use of this abbreviated form of the WAIS would be most efficient and valid and should be subject to experimental test.

B. Procedure

The object was to see how accurately the subtests that correlated most highly with the three factors revealed by Cohen could predict the Full Scale WAIS *IQ* of trainees. The subtests mentioned appeared to be the best measures of the individual factors; however, we needed to be sure that the Full Scale *IQ* obtained (prorated) would be very close to that obtained when all 11 subtests were administered.

Two approaches were used to test the predictive ability of the particular subtests: (*a*) prorating the Full Scale *IQ* on the basis of the six subtests and for the three factors; and (*b*) developing a regression equation which, given any score for the three factors, would predict the most likely Full Scale *IQ* score.

The subjects used in this study represented an accurate, cross-section of the average basic trainee population referred to the MHCD during the period of July, 1963 through July, 1965. There were 91 trainees distributed as follows according to diagnosis: Mild Mental Deficiency, 25 per cent; Character and Behavior Disorder, 35 per cent; Neurotic, 20 per cent; Psychotic, 4 per cent; Organic, 6 per cent; and Sexual Deviation, 10 per cent. Each *S* had been referred to the MHCD through normal command channels and the population used in this study was randomly selected. Eighty trainees were Caucasian and 11 were Negro. All of the trainees were in their first to eighth week of basic combat training at Fort Jackson. The mean age and education of the group were 20 and 10, respectively.

Each trainee was administered the entire WAIS (11 subtests) by a trained psychology technician judged to be competent to administer and score the WAIS by the author.

The prorating of the Full Scale *IQ* score was obtained by getting the average of the weighted scores for each of the six subtests used in obtaining the three factors and multiplying by 11. For example, if the weighted score average was 10, the Full Scale weighted score would be 110 and the resulting *IQ* equivalent could be derived from the tables in the WAIS manual according to the age range of the subject. A complete calculation is illustrated in the Appendix.

C. Results

The mean prorated *IQ* score of all 91 subjects was 90.84, while the mean Full Scale *IQ* (using all 11 subtests) was 90.45. We thus see that by using the prorated method we obtained an essentially identical *IQ* score with six subtests as we did with 11.

The regression equation derived from the data obtained looked as follows:

$$X = .81\text{M} + .22\text{NVPO} + .12\text{VC} - 13.14.$$

It is felt that the above mentioned application would be a more efficient way to evaluate differential cognitive functioning and overall general intelligence than is presently in use, while preserving an accurate measure of Full Scale WAIS *IQ*.

D. Summary

The purpose of this study was to develop an abbreviated form of the WAIS. The rationale for the selection of the six subtests used (Information, Vocabulary, Block Design, Object Assembly, Arithmetic, Digit Span) was based on Cohen's factor analysis of the WAIS. Both a regression equation and a prorating method were developed which compared quite favorably with the criterion.

Appendix

The following reveals the scores made by a trainee and his Full Scale *IQ* obtained by both the prorated and regression equation methods.

Age = 19

Subtest scores (weighted):

Information	10
Vocabulary	12
Block Des.	8
Obj. Ass.	6
Arithmetic	8
Digit Span	10

Derivation of *IQ*s:

VC: Average I and V (11) and multiply by 5 and refer to table in WAIS manual that converts weighted verbal scores to a verbal *IQ* (55 = 98 *IQ*).

NVPO: Multiply the average of BD and OA (7) by 6 and refer to tables converting performance weighted scores to an *IQ* (43 = 92 *IQ*).

M: Multiply the average of A and DS (9) by 5 and refer to performance section of manual (45 = 94 *IQ*).

GI: Multiply 11 by the average of the weighted scores (9). Refer to tables for conversion of total weighted score to total *IQ* score.

The regression equation derived is as follows: $X = .81(94) + .22(92) + .12(98) - 13.14$.

The *IQ* derived from the equation is 95.00, while the *IQ* derived from prorating was 95. We thus see the comparability of the two approaches. By using either one of the two predictive methods described, the requests of the MHCD were met in less than one-half the time formerly required. Periodically, a spot check was carried out by administering all 11 subtests and seeing whether the six special subtests used were predicting as accurately as in the original sample. In all cases ($N = 50$) the predictions were accurate.

REFERENCE

1. COHEN, J. A factor-analytically based rationale for the WAIS. *J. Consult. Psychol.*, 1957, **21**, 451-457.

290 Madison Avenue
Morristown, New Jersey 07960

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 87-95.

AFFECTIVE CORRELATES OF READING COMPREHENSION*

The University of Georgia

WENDELL W. WEAVER, WILLIAM F. WHITE, AND
ALBERT J. KINGSTON, JR.

A. INTRODUCTION

Few students of language have denied that affect plays a part in the acquisition, retention, and comprehension of language. Rankin (5) reported that extroverts (who are characterized as having greater impulsivity than introverts) scored lower on precloze tests of reading comprehension than did introverts. Since this difference was not apparent on postcloze tests, Rankin concluded that the "erratic responses" of the extroverts on the precloze test were responsible for this result. Rankin also found that "manifest anxiety" affected cloze scores. The cloze tests were related to anxiety because high anxiety subjects were "threatened" by the "initial difficulty of the items" (4).

Kingston and White (1) and White, Kingston, and Weaver (8) investigated the relationship of judgments of the concept *self* to judgments of the concept *protagonist* (of a story) by means of the Osgood Semantic Differential (SD). *S*s were divided into high anxious and low anxious, using the 16-PF second-order "anxiety" factor scores. Among high anxious *S*s, predictions of the general semantic meaning of the protagonist (evaluation, activity, and potency) could be made as well as predictions of the *S*'s concept of *self* (evaluation and potency). Among low anxious *S*s, only self-concept variables (potency and activity) were predictable from personality factors.

In the same study, extroverts and introverts were identified on the criterion of the second-order factors of the 16-PF. *S*s with relatively high second-order factor scores on "extroversion" projected "tangibility" into protagonist's ratings and "activity" into their own self-concept measures. Introverted *S*s projected "anxiety" onto the protagonist as well as "evaluating" their own self-concepts.

Osgood (2, 3) postulates a kind of "meaning" related to connotation and proposes that the Semantic Differential (SD) indexes this meaning. The SD required the construction of a "semantic space" by having *S*s judge concepts on a scale composed of bipolar adjectives. There is nothing in this technique

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 24, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

which prevents concepts having widely varying denotative meanings from falling in the same semantic space, or, conversely, nothing to cause concepts of very similar denotation from falling in widely differing semantic spaces. Furthermore, there is no assurance that the use of this procedure exhausts the domain "connotation of language." Nevertheless, the Osgood instrument has exhibited the power to index underlying affective components of language and deserves careful consideration in a study of affect and cognition.

Measures of cognition are often categorized by intended function—e.g., achievement, intelligence—or by type of item—e.g., multiple-choice, essay. The cognitive tasks of this study are of the deleted work variety—commonly called cloze tests. Postcloze tests are administered after the reading of a passage from which the deletion is drawn and are thus similar to achievement tests. It is much more difficult to decide how precloze tests should be categorized. They depend upon general background knowledge of the *S* and upon his ability to utilize the redundancy inherent in language. They correlate relatively high (in the 70's) with standard test of intelligence (7).

B. HYPOTHESES

A number of predictions were made about the relationship of cognition (as measured by the cloze procedure) and affect (as measured by the 16-PF second-order factors and the judgments of *self* and *protagonist* concepts on the Semantic Differential). These were as follows:

1. Precloze tests, the 16-PF, and the SD-self will share a dimension definable as "anxiety."

2. Precloze tests will be less predictable by high anxiety *S*s than will postcloze tests.

3. Precloze tests, postcloze tests, the 16-PF, and the SD-*self* will share a dimension definable as introversion-extroversion.

4. Precloze tests will be less predictable by introverted *S*s than by extroverted *S*s.

5. The *protagonist,* as measured by the SD, will be better predicted by low anxious than by high anxious *S*s.

6. The *protagonist,* as measured by the SD, will be better predicted by the introverted than by extroverted *S*s.

7. Precloze tests will be more closely related to sentences of the story which reflect affect: i.e., which contain "affect-ladened" words.

C. METHOD

1. *Subjects*

The *S*s were 80 senior division, undergraduate, female students enrolled in eight sections of an educational psychology course.

2. *Procedure*

Tests were given in the following order: precloze, SD-*self,* 16-PF, postcloze, SD-*protagonist.* *S*s read the passage, "I Starved for Science" (6), just before taking the postcloze and SD-protagonist tests. All *S*s were tested during the same two-week period. Three testing sessions were used. *S*s were told that the experiment was designed to study language, not to appraise or study their personalities. An attempt to motivate *S*s more fully was made by informing them that the results of the experimentation would be discussed with them.

3. *Analysis*

The 16-PF was scored in the routine manner and the raw scores factor analyzed. Six second-order factors appeared in the analysis. Although the components varied somewhat from Cattell's findings, it was felt that the analysis was close enough to Cattell's to use his designation in naming the factors.

The Semantic Differential was factor analyzed separately for *self* and *protagonist concepts.* Three factors were found in both analyses. These were identified as being similar in structure to Osgood's "activity," "potency," and "evaluation" factors. The factor scores from these analyses were used throughout the study.

D. RESULTS

The first four hypotheses were not supported by the data. No marked relationships were observed among precloze tests, 16-PF second-order factors, and the Semantic Differential within the same system. Variable interactions were noted when a principle components analysis was applied to the data (see Table 1).

Factor II in the analysis of the principle components was defined by the SD-self activity factor and introversion-extroversion second-order factor of the 16-PF. In an attempt to determine the relationships of the three SD-self factors and the six 16-PF second-order factors, a canonical correlation was obtained between the two sets of variables ($R = .60$, $p < .01$). While the SD-self activity coefficient was .97, the introversion-extroversion (16-PF)

TABLE 1
PRINCIPLE COMPONENTS FACTOR ANALYSIS OF PRECLOZE, POSTCLOZE, 16-PF, SD-SELF, AND SD-PROTAGONIST

Variable	Factor loadings											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	h^2
Precloze												
Struc-concept 1								57				.54
Struc-concept 2			48						40			.65
Lex-concept 1									80			.71
Lex-concept 2						76						.75
Struc-affect 1			76									.69
Struc-affect 2							56					.60
Lex-affect 1									67			.60
Lex-affect 2			68									.64
Postcloze												
Struc-concept 1	45										54	.72
Struc-concept 2	47											.67
Lex-concept 1	66											.68
Lex-concept 2	76											.75
Struc-affect 1	39			45							50	.79
Struc-affect 2	75											.61
Lex-affect 1	67											.63
Lex-affect 2	53							47				.72
Semantic Differential												
Self-activity		79										.70
Self-potency							78					.73
Self-evaluation					39							.70
Protag-potency										57		.61
Protag-evaluation											—73	.67
Protag-activity										—74		.78
16-PF												
Factor I					—89							.83
Factor II		—83										.74
Factor III				40		50						.76
Factor IV								77				.75
Factor V					40	—64						.72
Factor VI				82								.76

Note: Loadings below .39 are not considered in the analysis.

factor was —.81. The remaining coefficients were relatively low, ranging from .05 to .37.

Hypotheses 5 and 7 were tested by a canonical correlational analysis between the three SD-protagonist factors and the 16-PF second-order factors. One of the canonical correlations ($R = .41$) approached a significant level ($p < .10$), but no correlation was significant at the chosen .05 level of confidence.

Hypothesis 7 was not sustained in the form in which it was stated. There were differential effects of cloze results from sentences chosen to sample "affect."

To explore the data from promising hypotheses, other analyses were made. A step-wise regression was applied to cloze test responses and SD-protagonist measures. Results are found in Tables 2 and 3.

The activity factor, which tends to dominate all the relationships between SD-self and the 16-PF, exhibits a tenuous correlation to the postlex affect test

TABLE 2

STEP-WISE REGRESSION ANALYSIS OF THE ACTIVITY (SD-PROTAGONIST) FACTOR AS A DEPENDENT VARIABLE; CLOZE TESTS AS INDEPENDENT VARIABLES

Order of cloze variable entering	Cumulative multiple R	Cumulative multiple R-square
Postlex concept	.15	.02
Postlex affect	.28	.08*
Prestruc concept	.29	.09
Prelex concept	.31	.10
Prelex affect	.32	.10
Prestruc affect	.32	.10
Poststruc affect	.32	.10
Poststruc concept	.32	.10

* Significant at .05 level.

TABLE 3

STEP-WISE REGRESSION ANALYSIS OF THE EVALUATION (SD-PROTAGONIST) FACTOR, AS A DEPENDENT VARIABLE; CLOZE TESTS AS INDEPENDENT VARIABLES

Order of cloze variable entering	Cumulative multiple R	Cumulative multiple R-square
Poststruc concept	.29	.08*
Postlex concept	.33	.11*
Prestruc affect	.34	.12
Poststruc affect	.35	.12
Prelex affect	.36	.13
Prestruc concept	.36	.13
Postlex affect	.37	.13
Prelex concept	.37	.14

* Significant at .05 level.

(Table 2). In Table 3 two postcloze tests, which were deletions of sentences selected because of an apparent absence of affect meaning, were related to the SD-protagonist, evaluation factor.

Relationships between cloze tests and SD-self measurements appear to be more predictable than SD-protagonist and cloze test variables (see Tables 4 and 5).

TABLE 4
STEP-WISE REGRESSION ANALYSIS OF THE FACTOR ACTIVITY (SD-SELF) FACTOR, AS A DEPENDENT VARIABLE; CLOZE TESTS AS INDEPENDENT VARIABLES

Order of cloze variable entering	Cumulative multiple R	Cumulative multiple R-square
Poststruc concept	.26	.07*
Prestruc affect	.34	.11*
Postlex concept	.39	.15*
Prestruc concept	.42	.17*
Postlex affect	.44	.18*
Poststruc affect	.44	.19*
Prelex concept	.44	.20*

* Significant at .05 level.

TABLE 5
STEP-WISE REGRESSION ANALYSIS OF THE POTENCY (SD-SELF) FACTOR, AS A DEPENDENT VARIABLE; CLOZE TESTS AS INDEPENDENT VARIABLES

Order of cloze variable entering	Cumulative multiple R	Cumulative multiple R-square
Prestruc concept	.19	.04
Poststruc affect	.31	.09*
Prelex affect	.38	.14*
Prestruc affect	.39	.15*
Postlex affect	.40	.16*
Prelex concept	.41	.17

* Significant at .05 level.

The activity factor of the SD-self dimension is predictable from the entire battery of cloze tests in the analysis. Potency appears to relate selectively to those cloze tests which were constructed from sentences introspectively judged to contain "affect words."

A step-wise regression analysis was applied to the second-order factor scores of the 16-PF as dependent variables and the cloze test scores as independent variables.

Neither anxiety nor extroversion-introversion were observed to have any relationship to cloze tests, yet the second-order factors III and IV predicted cloze test scores. A single, well-defined postcloze test factor was identified, which included all postcloze test conditions: structural, lexical, concept-affect, and subtests 1 or 2.

The Semantic Differential ratings on the concepts of *self* and *protagonist* did not share significant variance.

E. DISCUSSION

There was no manipulation of the data which indicated that a general affective factor was operative in all the instruments used. In no analysis did the cloze tests, the 16-PF second-order factors, and the Semantic Differential appear in one system of relationships. There were, however, binary relationships between a number of pairs. These findings seem to indicate that affect does not readily appear as a general component underlying all types of affective and cognitive measurement. Valid measurements of affect seem to require specification of a relatively small domain. Even the Semantic Differential, which usually is taken to mean all things to measurements affect-wise, does not exhibit generality in this study. Rather it appears as a specific affect measure related to specific cognitive conditions.

The most stable binary relationship found was between the cloze tests and the SD-*self* "activity" factor. Perhaps a type of "temperament" variable may be involved here which allows highly active *S*s to attend more effectively to long stretches of rather boring material.

The "potency" factor of the SD-self was predicted by cloze tests of "affect." The "affect" cloze was constructed, introspectively, to contrast with "conceptual" cloze. The affect cloze sentences were of the type "Some of the men began to get bald," and "We let little doubts and worries get the best of us." So-called "conceptual" sentences were of the type "This fat was used in cooking," and "Some postponed part of their college work." When cloze deletions were carried out on this kind of material, all "affect" tests—pre and post, structural and lexical—were involved in the prediction of "potency." "Conceptual" tests were not involved. The expectation was that *S*s would be more inhibited on "affect" tests than on "conceptual" tests and, thus, performance would suffer. The opposite seemed to be true. *S*s scored much better on "affect" tests than on "conceptual" tests. Again an explanation may be that the so-called "conceptual" tests were just too dull to keep attention going until the tasks were completed. A check on incomplete items however revealed no more items were omitted on the conceptual cloze than on the affect cloze.

When we turn to the SD-protagonist, the first impressive result was that no precloze test predicts the Osgood factors for the protagonist. Since at this point in the study the *S*s have not read the passage in its integrated form, and have seen only randomly ordered sentences sampled from the passage, containing 20 per cent deletion of words, one would be hard-pressed to explain the relationships if they appeared. This result should also be considered in light of the prior finding with the SD-self, where pretests are as effective as posttests in predicting. Success with pretests in relation to the SD-self, then, are convincingly separated from success with pretest in relation to the SD-protagonist. One is encouraged to interpret the SD-self—cloze result as depending upon subject variables regarding his own person, while interpreting the SD-protagonist—cloze results as depending upon subject variables when the subject is indeed considering the character of the story which the experimenters presented.

No relationships were found between the 16-PF "anxiety" and "extroversion-introversion" factors and the cloze tasks. Cloze tests were predicted, however, by the second-order factors which Cattell calls "tough poise" and "independence." All cloze tests supplied variance positively related to the "tough poise" factor. In the case of the "independence" factor, the variance was supplied by postcloze tests only, and the high postcloze scores were negatively related to the positive end pole of the "independence" factor. This might be explained as a "rigidity" effect, with more dependent *S*s tightly bound to doing the task "right," even though it might not be intrinsically rewarding.

As Kingston and White (1) have noted before, there is a substantial relationship between the two major 16-PF factors and SD measures. For this particular sample the relationship between the activity dimensions of the SD-self and the extroversion-introversion dimension of the 16-PF was particularly strong. The relationship between the 16-PF "anxiety" and "extroversion-introversion" second-order factors and the SD-protagonist measure has been consistent in other studies of this nature and it is probable that only the limited *N* of *S*s employed in this study may underlie the failure of this relationship to attain statistical significance at the .05 level. It should not be surprising that relationships between a "personal" anxiety and a "personal" extroversion-introversion are more weakly related to affective judgments of the "main character" of a story than are SD-self concepts.

As may be seen by studying the tables in this study, a relatively small amount of variance, on the order of 10 to 20 per cent, is attributable to affective influences on cognition. This implies that errors of measurement in the instruments used to examine these relationships must be reduced as much as possible, or else the error variance will easily obscure the treatment effects. A number of cloze categorizations were combined in this study in order to insure higher reliability.

F. SUMMARY

A study was performed to examine the relationship between the cloze procedure (as a measure of reading comprehension) and affect measures (the Semantic Differential and the 16-PF). The three measures did not appear in one system of relationships. A number of binary relationships between the cloze procedure and the affect measures did appear, however.

It was concluded that the affect component in cognition is more specific than is generally assumed. Further studies are underway examining this specificity in detail.

REFERENCES

1. KINGSTON, A. J., JR., & WHITE, W. F. The relationship of reader's personality components to semantic meanings perceived in the protagonist of a reading selection. *Reading Res. Quart.*, 1967, **2**, 107-116.
2. OSGOOD, C. E. Studies on the generality of affective meaning systems. *Amer. Psychol.*, 1962, **17**, 10-28.
3. OSGOOD, C. E., *et al.* Analysis of the connotative meanings of a variety of human values as expressed by American college students. *J. Abn. & Soc. Psychol.*, 1961, **62**, 62-73.
4. RANKIN, E. G., JR. An evaluation of the cloze procedure as a technique for measuring reading comprehension. Unpublished Doctoral dissertation, University of Michigan, Ann Arbor, Michigan, 1957.
5. ———. Reading test reliability and validity as a function of introversion-extroversion. *J. Devel. Reading*, 1963, **6**, 106-117.
6. SIMPSON, E. A. Better Reading Book. Chicago, Ill.: Sci. Res. Assoc., 1962.
7. TAYLOR, W. L. Cloze readability scores as indices of individual differences in comprehension and aptitude. *J. Appl. Psychol.*, 1957, **41**, 19-26.
8. WHITE, W. F., KINGSTON, A., JR., & WEAVER, W. W. Affective dimensions in the connotative meaning in reading. *J. of Psychol.*, 1967, **67**, 227-234.

Department of Educational Psychology
The University of Georgia
Athens, Georgia 30601

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 97-116.

AN ANALYSIS OF THE USE OF ANALOGIES AND METAPHORS IN SCIENCE*

New School for Social Research

ROY DREISTADT

A. INTRODUCTION

It has frequently been pointed out by psychologists who have studied creativity that analogies and metaphors are often used by scientists in making their discoveries. The origin of this conclusion is the fact that many scientists themselves have said that they obtained their insight or illumination with the aid of an analogy. However, no comprehensive attempt has so far been made to study the use of analogies in science.

This is a study of the use of analogy and metaphor in science and is based on representative examples from all the major sciences—viz., astronomy, physics, chemistry, biology, psychology, mathematics, and computer engineering. A metaphor will be defined as an analogy expressed in verbal form. An analogy is in the form of sensory, usually visual, imagery. As will be seen, this research will show that analogies (and metaphors) can be simultaneously classified according to the type of experience in which they originate (Objective Analogies or Personal Analogies), the uniqueness or originality of the analogy used (elaborating an analogy already in use in a science, finding an analogy, and both finding and elaborating an analogy), and the ways that analogies aid in solving problems (analytic or synthetic). After this, additional observations will be stated, problems discussed, and a theory of creativity in science and insight will be postulated.

B. OBJECTIVE AND PERSONAL ANALOGIES

This classification is according to the type of origin analogies have. *Objective Analogies* are derived from external perceptual experiences and do not include the person himself as part of the analogy. These are the most numerous and come in great variety.

In the field of mathematics, Henri Poincare (17, p. 35) wrote:

> . . . the mathematical facts worthy of being studied are those which, by their analogy with other facts, are capable of leading us to the knowl-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 25, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

> edge of a mathematical law just as experimental facts lead us to the knowledge of a physical law. They are those which reveal to us unsuspected kinship between other facts, long known, but wrongly believed to be strangers to one another.

Another great mathematician, Charles Hermite, observed that other sciences, especially biology, may be very useful to study even for mathematicians because hidden and eventually fruitful analogies may appear between processes in both kinds of study (6). Hermite's example also shows that a person's interests, such as reading in other fields, often are a source of analogies. In many cases this interest in other fields is sufficient to cause the person actually to work in them, too.

In astronomy the discovery that the "luminous band around the heavens" called the Milky Way is actually a galaxy or island universe in which we ourselves are located was made by the analogy of a giant pinwheel or disk, which alone can explain why the Milky Way looks to us as it does. Sir William Herschel, who made this discovery, thought our sun was at the center of the Milky Way, but Harlow Shapley and others later showed that our sun is one of the stars in its periphery.

Sir Isaac Newton's famous analogy is mentioned by Voltaire who said he heard it from Newton's cousin who was supposed to have heard it from Newton. Newton related that the occasion of his discovery of the law of universal gravitation was the fall of an apple. This led to the thought that it was the falling of the moon toward the earth—because of the earth's gravitational force—that kept the moon in its path around the earth. The analogy compares the earth and moon with the earth and the apple.

That Newton often used analogies can be seen in his paper, "A new theory of light and color" (14, p. 46), wherein he tries (unsuccessfully) to use an analogy of a tennis ball curving when it is hit obliquely by a racket. He writes:

> Then I began to suspect, whether the Rays, after their trajection through the Prisme, did not move in curve lines, and according to their more or less curvity tend to divers parts of the wall. And it increased my suspition, when I remembered that I had often seen a Tennis Ball, struck with an oblique Racket, describe such a curve line.

However, Newton said that this analogy was incorrect and did not lead to his insight.

In arriving at a conception of the atom, the solar system was a useful analogy. The first reasonably adequate theory of the atom was the Bohr-Rutherford solar system atom.

Robert Oppenheimer (15, p. 129) states that ". . . analogy is indeed an indispensable and inevitable tool for scientific progress." He gave five examples of the use of analogy in physics. This is one example:

> Wave theory—it originated in the observation of regular, rhythmic changes in matter, waves on water, and was developed by an easily conducted physical exploration of sound waves, where there is a periodic change, a regular change in the density of air or other media . . . . At each point the scientists have tried to make a theory like earlier theories, light, like sound, as a material wave; matter waves like light waves, like a real, physical wave; and in each case it has been found that one had to widen the framework a little, and find the disanalogy which enabled one to preserve what was right about the analogy (15, p. 131).

James Clerk Maxwell (12) first stated his electromagnetic theory, which is a unified theory of electrical and magnetic events, in his article "On physical lines of force." He constructed an elaborate model to account not only for electrostatic effects (as he did in an earlier model) but also for magnetic attraction and electromagnetic induction. In the new model, "molecular vortices" rotating in space were the agents that produced magnetic fields. He pictured a molecular vortex as a slender cylinder which rotates around the lines of magnetic force. The velocity of rotation depends on the intensity of the magnetic force. Two mechanical effects are associated with these cylinders: tension in the direction of the lines of force, and lateral pressure arising from the centrifugal force produced by the rotating cylinders. Together these effects mechanically reproduce magnetic phenomena, since magnetism is a force exerted both along the axis and outward from the axis. He imagined a uniform magnetic field consisting of a portion of space filled with cylinders rotating at the same velocity and in the same direction "about axes nearly parallel." The cylinders can turn in the same direction because they do not make direct contact with each other but have rows of small spheres, like layers of ball bearings, between them which act as gears. He thought of the spheres or ball bearings as particles of electricity. In an unchanging magnetic field, all the cylinders rotate at the same constant rate. The little rotating spheres keep their position—so there is no flow of particles and no electric current. Maxwell pictured a change in magnetic force as a speeding up of rotation of the cylinders. As each cylinder is speeded up, it transmits the change in velocity to its neighbors. Since a cylinder now rotates at a slightly different speed from its neighbors, the spheres between them are torn from their positions by a kind of shearing action. This movement of the spheres is an electric current. In this way Maxwell explained how magnetic changes produce electric currents. He then explained how a change in electric force produces magnetism. If the spheres and cylinders are at rest and a force is applied to the spheres of electricity, causing them to move, the cylinders of magnetism with which they are in contact will begin to rotate, thereby producing a magnetic force. Furthermore, the cylinders will rotate in the direction perpendicular to the motion of the spheres, thus bearing out the observation that a magnetic field acts at right angles to the flow of a current.

The analogy he started with (and then elaborated) must have been between rotating cylinders in contact with spherical ball bearings that therefore rotate in the opposite direction, and his magnetic field producing vortices and particles of electricity.

Niels Bohr has said (2) that the idea of complementarity in physics came to him in the metaphoric form: that one could not know somebody simultaneously in the light of love and in the light of justice. This metaphor is from a theme found in Biblical and other literature.

Albert Einstein's answer to a questionnaire which Hadamard (6) sent to him was partly as follows:

> (A) The words or the language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The psychical entities which seem to serve as elements in thought are signs and more or less clear images which can be "voluntarily" reproduced and combined . . . .
>
> (B) The above mentioned elements are, in my case, of visual and some of muscular types . . . .
>
> (C) According to what has been said, the play with the mentioned elements is aimed to be analogous to certain logical connections one is searching for (6, p. 142).

One of many examples of Einstein using an analogy (1) is the following. He pictured a high building and inside it an elevator that was falling freely. Inside the elevator physicists, not aware that their ride might end in disaster, are performing experiments. When they take coins and keys out of their pockets and let go of them they remain in mid-air because they are falling along with the elevator and the men, at the same rate. They may explain these peculiar events by supposing that they have been transported outside the gravitational field of the earth and are somewhere in empty space where everything obeys Newton's Law of Inertia and continues in its state of rest or uniformly in a straight line. The elevator has become an inertial system and there is no way for the men inside it to tell whether they are falling in a gravitational field or are floating in empty space free from all external forces. Einstein then shifts the scene. The elevator is now really in empty space and the elevator is being pulled by a cable at the top of the elevator at a constantly increasing velocity. This time the men notice that their feet press solidly against the floor and if they jump they do not float to the ceiling because the floor comes up beneath them. If they let go of objects, such as keys, they appear to "fall." If they toss objects across the elevator they do not move uniformly in a straight line but follow a parabolic curve with respect to the floor. Since the scientists have no idea that their car is climbing through interstellar space they may conclude that they are situated in a stationary room on earth and affected by the force of gravity. There is no way for them to tell whether they are at rest in a gravitational field or climbing with constant acceleration through outer space where there is no gravity.

These imaginary events were aimed to be analogous to real events and Einstein used them to derive his Principle of Equivalence of Gravitation and Inertia—that there is no way to distinguish motion produced by inertia (e.g., acceleration, recoil, centrifugal force) from motion produced by gravitation.

His original analogy here must have been riding on an elevator and the experience one has of levity which somewhat counteracts gravity when the elevator goes down and the feeling of increased gravity (heaviness) when the elevator goes up. From this he must have guessed that inertia and gravitation are equivalent and *elaborated* the analogy to demonstrate it better. If this is correct, it is only his elaborated analogy which is an objective analogy, while his basic analogy of his experience while riding in an elevator is actually a personal analogy. Later, when the author discusses personal analogies, we will see that Einstein, in one of his earliest uses of an analogy, used a personal one. Perhaps as he got older he tended to use more objective analogies, or at least to put them in that form for publication.

In chemistry, Dmitri Mendeleev (8) discovered the periodic law and constructed the periodic table in 1869, from which he even corrected some erroneous atomic weights of elements and predicted three new elements from gaps in his periodic table which were later discovered. At that time 63 elements had been discovered. He took 63 cards, wrote the names and properties of the elements on them, and pinned the cards on the walls of his laboratory. Then he carefully reexamined the data, sorting out the similar elements and pinning their cards next to each other on the wall. A striking relationship thus was made clear. He discovered that the properties "were periodic functions of their atomic weights" that repeated themselves periodically after each seven elements. It seems probable to this writer that, since he wrote the elements on separate cards, he used playing cards as an analogy. The fact that playing cards follow each other serially and their suits repeat makes it a fairly good analogy.

He thus may already have been on the lookout for the periodic pattern even when he had only noted a few similarities among the elements.

John A. R. Newlands (8), an English chemist, anticipated by about three years the basic idea of the periodic law. In 1866 John Newlands read a paper before the English Chemical Society in which he compared the arrangement of the elements to the keyboard of a piano with its 88 notes divided into periods or octaves of eight. He said that the elements should be divided into octaves because each eighth element starting from a given one is in this arrangement a kind of repetition of the first, like the eighth note of an octave in music. He called this the law of octaves in chemistry. His use of this analogy was met with ridicule and almost ended his career as a scientist.

A somewhat different example of the use of analogy in chemistry is that of Friedrich August Kekule who discovered the "benzene ring" (usually represented as a hexagon), and developed his "closed-chain" or "ring" theory which showed the importance of molecular structure in organic chemistry. About the occasion of his discovery Kekule wrote (10, p. 118):

> I turned my chair to the fire and dozed. Again the atoms were gambolling before my eyes. This time the smaller groups kept modestly in the background. My mental eye, rendered more acute by visions of this kind, could now distinguish larger structures, of manifold conformation; long rows, sometimes more closely fitted together, all twining and twisting in snakelike motion. But look! What was that? One of the snakes had seized hold of its own tail, and the form whirled mockingly before my eyes. As if by a flash of lightning I awoke.

In this unusual example the analogy was seen in a dream or hypnogogic hallucination. His dream images were very likely influenced by seeing the smoke swirling around from the fire before he dozed. Smoke sometimes makes ringlike shapes or vaguely suggests them, just as Rorschach inkblots suggest things.

In biology Charles Darwin used "the living tree metaphor." This is the chief metaphor in *The Origin of Species* (7, p. 32):

> The affinities of all the beings of the same class have sometimes been represented by a great tree. I believe this simile largely speaks the truth. The green and budding twigs may represent existing species; and those produced during former years may represent the long succession of extinct species. At each period of growth all the growing twigs have tried to branch out on all sides, and to overtop and kill the surrounding twigs and branches, in the same manner as species and groups of species have at all times overmastered other species in the great battle for life. The limbs divided into great branches, and these into lesser and lesser branches, were themselves once, when the tree was young, budding twigs, and this connection of the former and present buds by ramifying

> branches may well represent the classification of all extinct and living species in groups subordinate to groups. Of the many twigs which flourished when the tree was a mere bush, only two or three, now grown into great branches, yet survive and bear the other branches; so with species which lived during longpast geological periods, very few have left living and modified descendants. From the first growth of the tree, many a limb and branch has decayed and dropped off; and these fallen branches of various sizes may represent those whole orders, families, and genera which have now no living representatives, and which are known to us only in a fossil state. As we here and there see a thin straggling branch springing from a fork low down in a tree, and which by some chance has been favored and is still alive on its summit, so we occasionally see an animal like the Ornithe-rhynchus or Lepidosiren, which in some small degree connects by its affinities two large branches of life, and which has apparently been saved from fatal competition by having inhabited a protected station. As buds give rise by growth to fresh buds, and these, if vigorous, branch out and overtop on all sides many a feebler branch, so by generation I believe it has been with the great Tree of Life, which fills with its dead and broken branches the crust of the earth, and covers the surface with its everbranching and beautiful ramifications.

In the first sentence of this interesting and elaborate metaphor one reads that the tree metaphor had already been used by biologists. Darwin's insight consisted in seeing more in the metaphor than they did.

There are a few other important analogies used by biologists. Hermann von Helmholz developed his theory of hearing on the analogy of musical instruments, such as the harp or the piano. His theory was that the fibers of the basilar membrane resonate to external frequencies somewhat like wires in a harp or piano.

Elie Metchnikoff, the Russian biologist, was observing the life of the mobile cells in the transparent larvae of starfish with his microscope, and idly threw a few rose thorns near them. The thorns were quickly surrounded by the larvae and dissolved inside their transparent bodies—they had been gobbled up and digested. This reminded him of what happens when a human finger is infected by a splinter—and he thought that the pus which surrounds it attacks and tries to digest the intruder like the starfish larvae. By this analogy, Metchnikoff discovered man's main defense mechanism against invading microbes: the phagocytes which are a population of cell-eating white blood corpuscles (10).

Paul Ehrlich, as a university student, spent much of his time staining almost everything. Just as certain dyes stain only some cells and not others, he was analogously looking for substances which would act as "magic bullets" and act only on their targets, specific disease microbes. He discovered salvarsan, an arsenic compound, and other compounds, and became the founder of modern chemotherapy (5).

Recent work on DNA and RNA uses metaphors—that its code is an *alphabet* spelled with four letters and the structure of DNA is sometimes likened to a spiral staircase.

In psychology, William McDougall used the hydraulic analogy in his drainage theory of the nervous system to explain inhibition, Pavlovian conditioning, and other events. He explained neural phenomena on the analogy of hydraulics—basically, less water comes out of a faucet one has turned on (inhibition) when someone else has the water turned on.

Modern behavior theory is based on analogy. When behavior theorists, such as C. L. Hull, E. C. Tolman, E. Guthrie, O. H. Mowrer, K. Spence, and Neal Miller, develop their theories mainly on the basis of experiments with lower animals (especially the white rat) and then use these findings to help them develop a psychological theory that applies also to man, they are working on the basis of analogy.

Wolfgang Kohler's theory of electrical fields in the brain is analogous to field theory in chemistry and physics. In a less direct manner all gestalt theory is analogous to field concepts in physics. Kurt Lewin's behavioral field and topology theories are in a general way analogous to field theory in physics and to the mathematical branches of topology and non-Euclidian geometry.

One of Jean Piaget's basic analogies is from group theory in mathematics developed by Galois. Piaget (16, p. 40) gives evidence of his analogy when he writes about the conditions for the operations of intelligence: "These conditions are four in number in the case of 'groups' of a mathematical order, and five in the case of 'groupings' of a qualitative order." Then he lists (16, p. 42) the five conditions of a psychological grouping which is one more than for mathematical groups: I. combinativity, II. reversibility, III. associativity, IV. general operation of identity, and V. tautology or special identities.

Finally we come to Freud, who used many objective analogies. Freud suddenly got the idea of sublimation when he saw a cartoon which is a metaphorical illustration of sublimation:

> It happened while he was looking at a cartoon in a humorous periodical which showed the career of a young girl in two subsequent stages. In the first she was herding a flock of young geese with a stick, in the second she was shown as a governess directing a group of young girls with her parasol. The girls in the second picture were arranged exactly in the same groups as the goslings in the first (18, p. 98).

In his theory of dreams, the concept of censorship in dreams was arrived at by use of an analogy of a "watchman"; and the use of terms, such as cathexes, anticathexes, and hypercathexes, in developing his psychoanalytic theory are electrical metaphors. He also used picture language analogies when he wrote that dreams are in "a pictographic script" and that dreams "present no greater difficulties to their translators than do the ancient hieroglyphic scripts to those who seek to read them" (7, p. 336).

An interesting analogy Freud used (3, p. 298) to bring together his theories of neurosis, regression, and fixation was the following:

> If you think of a migrating people who have left large numbers at the stopping places on their way, you will see that the foremost will naturally fall back upon these positions when they are defeated or when they meet with an enemy too strong for them. And again, the more of their number they leave behind in their progress, the sooner will they be in danger of defeat.
>
> It is important for comprehension of the neuroses that you should keep in mind this relation between fixation and regression.

Freud also tended to explain psychological and anthropological events *by using one as an analogy for the other.* He used myths to understand dreams and explained myths like dreams. He regarded religion as a cultural compulsion neurosis and compulsion neurosis as a personal religion.

This *dual manner of using objective analogies* is also found useful by computer engineers and brain physiologists across their sciences. On the one hand it is supposed that the "faculties" built into computers are analogous to those of the brain, and on the other hand it is supposed that the brain functions analogously to a computer. So, while engineers have put "memory" into computers (by closed reverberating circuits), built chess-playing computers which can learn from their mistakes (by negative feedback to their circuits), and are even trying to build computers that think creatively, brain physiologists are trying to explain human and animal memory in terms of closed reverberating circuits with chemical changes, learning to adapt to the environment by negative feedback, and motivation as a re-establishing of a steady state through negative feedback. Sometimes positive feedback occurs in a computer whereby output and input keep increasing each other and getting further and further away from a steady state. Neurotic behavior is supposed to be analogous to positive feedback (21). The theories of D. O. Hebb are strongly influenced by computer analogies.

Norbert Wiener, Gray Walter, and W. S. McCulloch also use the analogy of television to explain some brain processes. The alpha wave, which is the normal rhythm of the visual cortex, is supposed to serve as a scanning and carrier wave which picks up "the picture" coming into the brain and codes it somewhat as TV signals are coded, and then the signals are carried to different parts of the cortex for decoding. Any of the cortical areas can act as TV receivers by decoding the signals carried on alpha waves. This theory is supposed to solve the problems of equipotentiality and recovery of brain functions (11, 20).

Personal Analogies are of two types—normal personal analogies and pathological personal analogies (while the person is in some kind of pathological state). In all personal analogies the person himself is included within the analogy and makes up part of the analogy. To show the use of a *Normal Personal Analogy,* we have examples from Archimedes and one from Einstein.

King Hiero put the question to Archimedes whether a crown made for him and purporting to be made of gold did not actually contain some silver. Archimedes is said to have been puzzled until one day, as he stepped into a bath and observed the water running over, it occurred to him that the proportion of alloy could be measured by putting the crown and equal weights of gold and of silver separately into a pan of water and noting the differences of overflow. When this happy thought struck him he was so overjoyed that he ran home without his clothes, shouting *"eurika, eurika"* ("I have found it, I have found it"). Archimedes discovered the technique of measuring the gold content of the crown by displacement of water, by identifying his own body displacing water with that of the crown (this identification was probably easy for him to make, since he was a relative of the king). This is the analogy. The details of getting equal weights of gold and silver to use as standards came later (10; 19, p. 419). This is an example of a normal personal analogy because Archimedes himself is included in and is a part of the analogy. Archimedes seems to have been unable to solve the problem earlier because the water did not overflow before when he stepped into the bath. Until this occurred, there actually was no analogy!

From this same analogy Archimedes formulated the theory of specific gravity, still called the Principle of Archimedes, that "a body immersed in fluid loses as much in weight as the weight of an equal volume of the fluid."

Archimedes regarded his discovery of the relationship between the surface and volume of a sphere and its circumscribing cylinder as his greatest discovery and had a sphere inscribed in a cylinder marked on his tomb. To make this discovery he built a cylinder of which the height and diameter were the same and put a tightly fitting sphere into it. He filled the cylinder with water and then put the sphere into it. By comparing the *overflow of water* with the original amount of water in the cylinder, he discovered that the volume of an inscribed sphere is equal to two-thirds of the volume of the enclosing cylinder (5). It can be seen that he solved this problem by using the same analogy again!

Einstein (10, 13) visualized himself as a passenger riding on a ray of light and holding a mirror in front of him. This would cause no image of himself, for the light and the mirror are traveling in the same direction at the same velocity and, since the mirror is a little ahead, the light can never catch up to the mirror and there can be no reflection. Like a fictional vampire, he would cause no image. But this applies only to *his* mirror. A stationary observer, also equipped with a mirror and watching the rider flash by, will catch the rider's image. This fantasy shows that optical events are relative, and, since it is intended to be analogous to real events, Einstein makes his deductions from the fantasy situation and assumes that they explain the real world.

Pathological Personal Analogies have sometimes been used to solve problems in science. There are three good examples of this type of analogy which the author will discuss.

Geronimo Cardano, or Jerome Cardan (5, 6), was the most outstanding mathematician of his time. "It could be naturally expected that the discovery of imaginaries (imaginary numbers in mathematics) which seems nearer to madness than to logic and which, in fact, illuminated the whole mathematical science, would come from a man . . . who from childhood suffered from fantastic hallucinations to such an extent that he was chosen by Lombroso as a typical example in the chapter 'Genius and insanity' of his book *The Man of Genius*" (6, p. 135). It appears to this writer that Cardan's own tendency to regard his imagination as real, as perception, when he had hallucinations (and the fact that he probably somewhat was aware of this, since he was also a physician) caused him *analogously* to treat "imaginary" numbers in equations the same as real numbers. He did in mathematics analogously what he did in his everyday life.

Alfred Russel Wallace wrote in his autobiography (10, p. 142) that he was "lying muffled in blankets in the cold fit of a severe attack of intermittent fever at Ternate" (an island near New Guinea) when he suddenly remembered Malthus' essay on population which he had read "twelve or more years earlier." He continued: "It suddenly flashed upon me that this self-action process (i.e., the struggle for existence) would necessarily improve the race, because in every generation the inferior would inevitably be killed off and the superior would remain—that is, the fittest would survive. The more I thought it over the more I became convinced that I had at length found the long-sought-for law of nature that solved the problem of the origin of species." Koestler apparently thinks this is self-explanatory, but the writer thinks that one cannot see anything in this story that tells why Wallace should have gotten the sudden flash of insight of the "struggle for existence" and "survival of the fittest" until we say that Wallace—in his own illness—noted that *he* was struggling for existence and wondering whether he was fit enough to survive, and then had the insight that his own experience was an analogy of what might be the answer to the problem of what causes evolution. If this is correct, it is the type of analogy that includes the person himself, just as is the case with the analogy of Archimedes' Eurika experience—except that in Archimedes' case he was in a normal state of health. Wallace used a pathological personal analogy in which he identified himself (struggling to survive from his illness) with the species.

This interpretation, that Wallace used this particular personal analogy when he discovered the theory of evolution, is supported by the biographical fact (4) that as he got older he was also so much concerned about his personal survival after death that he turned his interest to spiritualism!

Actually, this is the only *comparative example* of the *same* discovery (evolution) being made by the use of an *objective analogy* (Darwin) and a *personal analogy* (Wallace).

Gustav Theodor Fechner, before founding psychophysics, had developed an illness induced by staring at the sun for long periods of time in order to study after-images, and during this illness he could only live in a partially darkened room. This writer thinks that it is reasonable to suppose that this partially darkened room in which he lived is a kind of *threshold* between daylight and darkness. As Fechner gradually recovered he could come out into the daylight: that is, above the threshold. Each day is an equal interval and on each day he could comfortably experience a geometrically increased intensity of light. His own experience would then be analogous to the logarithmic relationship between sensation and stimulus, which he expressed in Fechner's Law.

This analogy is like Wallace's which includes the person himself in the analogy in the state of his temporary pathological condition. Not only does the pathological state, in these cases, turn the person more to himself and fantasies about himself (introversion), but by analogy his state helps to provide the content of his discoveries, too.

In terms of combined behavior theory and phenomenological-existential theory, an *objective analogy* is an external stimulus pattern of external objects or events (experienced as an external sensation) and a *personal analogy* is an external stimulus pattern of external objects or events (experienced as an external sensation) *plus* internal, proprioceptive, or external (e.g., visual) stimuli of some kind which the person has of himself (experienced as internal, proprioceptive, or external sensations of himself). The personal analogy of Archimedes, for example, was all of external visual stimuli or sensations—of the bath and water plus his body stepping into the water and making it overflow. This describes the kind of stimuli that make up the analogies.

C. Analogies that Are Elaborated, Discovered, and Both Discovered and Elaborated

Next we will see that, whether an analogy is classified as an objective or personal analogy (normal or pathological personal analogy), it can further be classified according to the uniqueness or originality of the analogy used. There are three types here—elaborating an analogy already in use in a science, discovering an analogy, and both discovering and elaborating an analogy. A good example of a discovery made by elaborating an analogy already in use in his science is Darwin. The tree analogy was used by others during and before him, but they only used its branching shape to indicate the relationship of living organisms. Darwin saw more in the tree analogy than the others did. He arrived at his theory of evolution *by elaborating the tree analogy into a living tree analogy*. This is an example of elaborating an objective analogy already in use in a science.

Examples of discovering an analogy are Newlands, who discovered the musical analogy of octaves to classify the chemical elements, and Archimedes, whose body caused some water to overflow while he was taking a bath and who used this as an analogy to solve the problem of the king's crown. Newlands discovered an objective analogy and Archimedes discovered a normal personal analogy.

Finally, as examples of both discovering an analogy and elaborating it, we can cite Maxwell's analogy and most of Einstein's analogies. It has already been pointed out from what kinds of experiences these analogies probably were discovered and then elaborated. Maxwell's analogy is an example of discovering and elaborating an objective analogy. Maxwell's analogy was an objective analogy from the beginning and stayed that way when he elaborated it. However, the analogies Einstein found and elaborated involved people (observers) *which are part of the analogies themselves* and so were originally personal analogies. Einstein himself was the person riding on the ray of light and in the elevator. When he found these personal analogies, he converted some of them into objective analogies containing other people (observers) as he elaborated them but left some (such as the analogy of himself riding on a ray of light) in the form of personal analogies when he elaborated them.

When one experiences a stimulus pattern which one perceives to be an analogy that can be used to solve the problem, one can assimilate memory traces to it (which is the occurrence of some degree of reintegration) to make the analogy-stimulus richer in its structure (pregnance), as in the case of Darwin who saw more in the tree analogy than others had seen. Maxwell's elaborated objective analogy is another example of this. In all these cases analogies are elaborated. To elaborate an analogy one accepts only memory traces that assimilate to the analogy *which one can use to help solve more of the problem.*

The further problem here is: how does one discover an analogy in the first place when one discovers a new analogy? We will return to this question later.

D. Analytic and Synthetic Analogies

Finally, we come to the third part of the classification of analogies—what analogies do to aid the scientist in solving the problem he is working on. Each analogy, regardless of how it has been classified in other respects, can also be classified in terms of how it aids in solving the problem.

An example of an *analytic analogy* was the development of the solar-system atom theory. In the case of the atom, the problem was to *differentiate the whole into the correct kind of parts and to determine how the parts function within the whole*—and the solar system analogy showed how to do this.

Some examples of synthetic analogies are the following. The analogy of a disk or giant pinwheel to describe the Milky Way tells us about the shape and spin of the Milky Way (the positions and movements of its stars) as well as why it appears as a band because of our position in it.

Freud's analogy of migrating people is a synthetic analogy which shows how fixation and regression are related to each other and to neuroses.

Newlands' musical octave analogy is a synthetic analogy which shows how to put the chemical elements into the correct classification.

The tree analogy, as used by biologists before Darwin, was a synthetic analogy that showed the relationships between all living organisms. Many biologists, including Darwin, worked to improve this classification. But Darwin's big question really was why the "tree" has its tree shape! What causes a real tree to grow in a certain way and what limits the growth of some parts of it, is what he used as his analogy to explain what factors caused the growth and destruction of species in the tree of life. He then listed the influences that limit the growth of parts of a tree and generalized what these influences have in common—the fittest survive against competition—and used this to explain evolution. He used the tree analogy as a temporal analytic analogy for finding the factor or cause that determines the evolution of the tree of life.

Whether one uses an analytic or synthetic analogy depends on the problem —whether one has the whole and tries to discover its parts or has some parts and tries to discover the whole to which they belong. There appear to be different kinds of analytic analogies—some show the nature of the parts and their functions within a whole (solar system analogy), some show what factor temporally caused the whole to be as it is (Darwin's living tree analogy), and there are cases where an analogy helped solve a problem by showing that there is a temporal sequence of parts needed to cause a whole. In chemistry compounds are often found made of chains of elements that have a particular sequence and have to be synthesized that way. Many experiences were or could have been used as analogies for this kind of problem. For example, before we can mail a letter we must write it, before we can write it we need paper, a pen, etc.

Synthetic analogies are also not all the same. The pinwheel analogy shows the shape of an existing whole (the Milky Way), but the musical octave analogy shows the classification of elements (which is a whole but not a whole that exists together).

E. A THEORY OF CREATIVITY IN SCIENCE AND INSIGHT

Analogies are used in science to show how parts are put together into wholes or how to differentiate the parts of wholes, so that problems are solved. This gives the insight. Kohler does not really tell us how the sudden reorganization and insight occur, as most behavior theorists have argued (9). There is more to it here than a simple closing of small gaps (closure) because whole sections are missing here. *This writer explains insight as occurring when one finds a stimulus pattern (the analogy) in which parts of the form or structure are like the structure of the problem-situation and the rest of the structure of this stimulus pattern (the analogy) indicates how to organize the unintegrated materials of the problem or how to reorganize the problem by putting the parts that are out of place into their correct place, or both, thereby completing the whole which is then the solution of the problem.* There is also the reverse of this, where the problem is in the form of a poorly differentiated whole and an analogy helps one to discover *how* it is actually differentiated, or to find a more correct differentiation of a whole than the one that has been attained thus far. This kind of situation exists regarding the work of physicists with the atom. In either case, when the problem is solved, we have then *ipso facto* also created good continuation, closed gaps, etc., because the analogy—by showing how to put the parts together to make a whole or by showing what kind of parts make up the whole, in such a way that solves the problem—thereby indicates *how* the gaps get closed and *what* good continuations get established. Only small gaps which classical gestalt psychology deals with can be closed by themselves; big gaps where large parts or even most parts are missing can only be closed with the aid of an analogy that tells us *how* the gaps are to be closed. This entire analysis is an important advance over some parts of the theory of Kohler and his disciples, and opens up new vistas in gestalt theory and the study of creativity.

How does one find an analogy in the first place? As more and more data are gathered by scientists and some similar parts are put together (or somewhat similar parts are placed near each other tentatively, partly by trial and error), the qualitative samples that make up the data that has been collected and partially assembled eventually begin to appear like the corresponding parts of an analogous situation to a highly creative person who then completes the arrangement of the data the way it is in the analogous situation. This is how the sciences develop. For example, when Newlands discovered the periodic law in chemistry (called by Newlands the law of octaves), 63 elements were already known and it had already been noted that some elements were similar to others, that some elements repeated the characteristics of others (to different degrees), and that the elements seemingly could be ordered (with many big gaps) according to their weights ranging from the lightest to the heaviest. (Weight can be used to order *all* the elements because the elements are in different degrees similar in weight, and weight is the characteristic *all* the elements have.) When they were thus ordered according to weight it was seen that the elements that are next to each other often have *other* similar characteristics and that some characteristics of the elements tend to repeat periodically—but all this appeared vague with many big gaps. Despite the big gaps with entire sections of elements missing, the classification of the elements thus far attained suggested to Newlands the analogy of the keyboard of a piano with its 88 notes divided into periods or octaves of eight. The problem is much like a jigsaw puzzle that has been partly assembled so that we see parts of the picture of the puzzle which we interpret by recalling that the parts we see look like parts of a sailboat we have seen. We use the sailboat of our memory-image (which never is exactly the same sailboat as in the puzzle-picture) as an analogy to aid us in putting the rest of the pieces of the puzzle together. This is how one discovers a synthetic analogy.

An example of discovering an analytic analogy occurs in the development of the first modern concept of the atom. By bombarding the atom, Rutherford found that there is something more solid in the center than in the rest of the atom, whereas particles in the periphery become readily detached in chemical reactions. Rutherford was beginning to obtain some vague details of the structure of the atom. This suggested the solar system as an analogy because it looks like the vague atom picture he was discovering but is more clear and distinct, just as a person looks at the nebulous Rorschach inkblot number 5 and it suggests a bat which looks like the inkblot but is more clear and distinct. The person who can give many original (O) responses on the Rorschach which also are wholes (W) with good form (F plus) would be the most intellectually creative.

This interpreting of incomplete data of problems according to their analogy to other stimulus patterns (or their memory traces) is but a special case of what we do in all perception—for not only do we fill in small gaps as the gestalt psychologists have shown that we do, but when large sections of the sensory input are vague or meager we use memory-images that are partly like the parts we have of the sensory input and fill in the rest according to memory. This is in Ames-like fashion. In this manner we bet on meager sensory input.

There is an old controversy concerning two different accounts of scientific investigation, one based on induction and the other on deduction. Briefly, the inductivist account states that hypotheses are arrived at by generalization from observed instances, and the hypothetico-deductive account states that the scientist begins with hypotheses and then tests the deducible consequences of the hypotheses by observation. The inductive account does not succeed in really telling us how hypotheses are arrived at from observed events, and the deductivist says scientific hypotheses have been arrived at by all sorts of methods. Now this problem can be largely resolved.

In scientific investigation one begins by making observations of different events and their frequency of occurrence, and the observations *begin* to make a pattern as one tries to classify similar events (fit them or group them together). One obtains *hypotheses* of *how* things are (what kind of pattern it is) or *why* events are that way *by comparing this emerging pattern of events thus far assembled with a different event whose pattern includes parts that are like the events that have already been assembled in the problem one is studying. This is the use of analogy; we use the pattern of these other events as models—analogies are models and models are hypotheses.* Finding the analogy that we can use to extend or complete the pattern of events we are studying (or to differentiate a vague whole event we are studying) *is obtaining and experiencing the insight.* Thus we begin with the inductive method by collecting data from experience and then find an analogy which is used as the hypothesis or generalization. A generalization is a hypothesis. Can one make a generalization that is not a hypothesis? or frame a hypothesis that is not a generalization? No. [Even to make the simplest hypothesis—that the sun will rise tomorrow—is based on the supposition that tomorrow will be analogous to today and days in the past. Such a hypothesis is a generalization. All more complex generalizations (hypotheses)—like Darwin's theory of evolution —are statements describing the one pattern that is occurring in various different events which have been found to be analogous to each other.] The hypothesis (model) obtained from analogy tells us how to put together more completely the observations already made which had been only slightly put together, or how to differentiate a whole event more clearly, *and* even what further events one should look for to complete the pattern of events like the analogy or to differentiate a whole event more like the analogy, to demonstrate or prove the correctness of the hypothesis (analogy). Here he uses the hypotheticodeductive method in making further observations to find these additional events and their frequency that are needed to "prove" or "demonstrate" the hypothesis.

F. Additional Observations and Problems for Further Study

The analogies described by a scientist may not always be the ones he used to solve the problem. He may have solved the problem by one analogy and then thought of a better analogy which he published. This problem is really only important in biographical studies; it does not affect a study of the classification of analogies and their role in insight in science, such as is here attempted.

Analogies decrease the amount of imagery necessary to solve a problem and so the whole idea is able to occur suddenly in a "flash" of illumination or insight. The analogy tells us *how* to think about the problem in order to be able to solve it.

Aristotle wrote in the Poetics that "metaphor, the perception of likeness in dissimilars, is the sign of genius."

Analogous discoveries made in science can often be stated in terms of the same mathematical equations. For example, Newton's law of gravitation, Coulomb's law of electric charges, and the law of attraction of magnetic poles are stated by equations that have the same form. (Einstein, in fact, tried to develop a complete unified field theory that would show why so many equations in physics have the same form.)

As to the possible birth and death of analogies some writers speak of, it should be said that analogies do not die. They are later used for illustrative purposes to popularize a science when writing for nonscientists and to help teach the sciences to students.

A personal analogy could also be called a subjective analogy. We can also say that objective and personal analogies are indirect and direct existential analogies, respectively.

Personal (or direct existential) analogies occur on a deeper more basic mental level where the objective and subjective merge—since this analogy consists of external events and the person himself, combined together in some way (both of which make up the personal analogy). Objective analogies consist only of the external events.

There may be a preference in the analogies used (in creative activity) depending on the style of work of a scientist, which is an expression of his style of life. Using personal and objective analogies may be somewhat related to introversion and extraversion, respectively. In this connection it must not be overlooked that illness makes one temporarily more introverted (libido or interest returns from objects to the ego) and therefore favors the use of personal analogies. This relates to what the author already said when he discussed personal analogies, but further research here should help us find what relationships there are between the personality of a scientist and the discoveries he makes.

G. SUMMARY

This article analyzes the part that analogies play in obtaining the insights or illuminations of great scientific discoveries. Many examples of analogies that have been used in all the sciences are then classified as objective and personal analogies (the latter include the person himself in the analogy and may be either normal or pathological). Beyond this, analogies are further classified in terms of whether they were already used in science but were elaborated further, or whether they were discovered, or whether they were both discovered and elaborated. Finally, the analogies are classified still further as being analytic or synthetic depending on the way they help to solve a scientific problem, and it is shown that in either case an analogy shows how incomplete data are to be completed to give the solution to a problem. After analogies have been used to make discoveries, the analogies are used to help popularize and help teach the sciences. Finally, a theory is presented of scientific problem-solving in science based on a new extended version of the gestalt theory of insight. It is also shown how this theory resolves the controversy as to whether the scientific method is inductive or hypothetico-deductive.

References

1. Barnett, L. The Universe and Dr. Einstein. New York: William Sloan, 1948.
2. Bruner, J. S. The conditions of creativity. In H. E. Gruber, G. Terrell, & M. Wertheimer (Eds.), *Contemporary Approaches to Creative Thinking*. New York: Atherton, 1964. Pp. 1-30.
3. Freud, S. A General Introduction to Psychoanalysis. New York: Garden City, 1943.
4. George, W. Biologist Philosopher: A Study of the Life and Writings of Alfred Russel Wallace. New York: Abelard-Schuman, 1964.
5. Greene, J. E. One Hundred Great Scientists. New York: Washington Square Press, 1964.
6. Hadamard, J. The Psychology of Invention in the Mathematical Field. Princeton, N.J.: Princeton Univ. Press, 1945.
7. Hyman, S. E. The Tangled Bank. New York: Atheneum, 1962.
8. Jaffe, B. Crucibles: The Study of Chemistry. New York: Simon & Schuster, 1957.
9. Kimble, G. A. Hilgard and Marquis' Conditioning and Learning. New York: Appleton-Century-Crofts, 1961.
10. Koestler, A. The Act of Creation. New York: Macmillan, 1964.
11. Morgan, C. T., & Stellar, E. Physiological Psychology. New York: McGraw-Hill, 1950.
12. Newman, J. R. James Clerk Maxwell. *Sci. Amer.*, 1955, **192**(6), 58-71.
13. ———. Einstein's great idea. In S. Rapport & H. Wright (Eds.), *Mathematics*. New York: Washington Square Press, 1964. Pp. 78-95.
14. Newton, I. A new theory of light and color. In W. Dennis (Ed.), *Readings in the History of Psychology*. New York: Appleton-Century-Crofts, 1948.
15. Oppenheimer, R. Analogy in science. *Amer. Psychol.*, 1956, **11**, 127-135.
16. Piaget, J. The Psychology of Intelligence. Patterson, N.J.: Littlefield, Adams, 1960.
17. Poincare, H. Mathematical creation. In B. Gheselin (Ed.), *The Creative Process*. Berkeley, Calif.: Univ. Calif. Press, 1952. Pp. 33-42.
18. Sachs, H. Master and Friend. London: Imago, 1946.
19. Stagner, R., & Karwoski, T. F. Psychology. New York: McGraw-Hill, 1952.
20. Walter, W. G. The Living Brain. New York: Norton, 1953.
21. Wiener, N. The Human Use of Human Beings: Cybernetics and Society. New York: Houghton Mifflin, 1950.

256 East 89th Street
New York, New York 10028

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 117-120.

PERSONALITY CORRELATES OF WITHDRAWAL FROM SMOKING*[1]

Michigan State University and the University of Maryland

ARTHUR RESNIKOFF, PAUL G. SCHAUBLE, AND ROBERT H. WOODY

A. INTRODUCTION

Much of the research dealing with cigarette smoking has been concerned with comparisons of smokers with nonsmokers in regard to variables relating to demography, personal history, and temperament (1, 2, 6, 7). There has been little attention to the personality characteristics of smokers who are able to stop smoking as opposed to smokers who attempt to stop, but cannot. This study investigates personality variables of persons who have a history of smoking, but were motivated to terminate the smoking habit; their motivation was evidenced by their voluntary enrollment in a public smoking-withdrawal clinic. A 224-item "smoking questionnaire" was derived from the Minnesota Multiphasic Personality Inventory (MMPI), based upon research with the full scale instrument (4, 5, 9). Included in the questionnaire are the *D, Hy, Pd, Ma, Si,* and *Ne* scales with the appropriate *K* corrections. It was hypothesized that those *S*s who were unable to withdraw from smoking would have significantly higher scores on all scales than those *S*s who were successful in withdrawing from smoking.

B. METHOD

From persons volunteering to enroll for treatment in a smoking withdrawal program, 17 males and 20 females were assigned to the group of patients treated and investigated in this study. The combined age range for both sexes was 19 to 75 years, mean age 40 years. The *S*s met once a week for four weeks. At the first three meetings, *S*s received an educational therapy session, physical examinations were given, and medications were distributed. The educational therapy and medication program have been described in detail in a previously published related study (3). The "smoking questionnaire" was administered at the first meeting, and measures of smoking status were compiled at the first and fourth meetings. Since the initial relationship between

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 26, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Appreciation is expressed to Dr. C. Ross, Dr. S. Graham, and Mr. B. Heller.

amount smoked and questionnaire scores might have an influence on withdrawal success (defined herein as complete withdrawal from smoking by the fourth meeting), the correlations between amount smoked prior to entering the program and questionnaire scores were determined, and no correlations were significant at the .05 level.

C. Results and Discussion

Table 1 presents the mean scores of the *S*s on the questionnaire scales as compared with mean scores and standard deviations for the MMPI norm group. On all but the *Ne* scale (where standardized norms are unavailable), the mean for the *S*s was higher than the mean for the norm group, with score differences on the *Hy* and *Pd* scales being greater than one standard deviation from the mean. The application of a chi square ($\chi^2 = 1.51$) revealed that there was no significant difference in withdrawal success between the sexes at the .05 level.

TABLE 1
Comparison of Volunteers with Norm Groups on the Scale Scores

Scale	$\overline{X}$ experimental male	$\overline{X}$ norm male and 1 SD + or —		$\overline{X}$ experimental female	$\overline{X}$ norm female and 1 SD + or —	
D	20.41	17.0	12—21	22.43	19.5	14—24
Hy	23.35*	16.5	11—22	26.05*	19.0	13—24
Pd	23.82*	19.0	15—23	23.43*	19.0	14—23
Ma	19.82	17.0	13—21	19.09	17.0	13—21
Si	26.65	25.0	15—33	32.48	25.0	15—33
Ne	7.59	Unstandardized		9.57	Unstandardized	

* Beyond + or — 1 *SD*.

The means and *t*-scores for the successful *versus* the unsuccessful groups on each of the six scales are presented in Table 2. While most differences were in the hypothesized direction (exceptions being the female scores on the *Ma* and *Pd* scales), only the *Si* scale scores achieved statistical significance, and this difference appears meaningful only for the female *S*s.

In Table 3, cutting-scores are presented which differentiate the successful and unsuccessful female *S*s on the *Si* scale. The *Si* scale has high predictive value for this sample, differentiating females on success of withdrawal with an 87.5 per cent degree of accuracy.

While differences in scores on the *D, Hy, Pd, Si,* and *Ne* scales were consistently higher for *S*s who were unsuccessful as compared with those who were successful in their attempt to withdraw from smoking, the only sig-

TABLE 2

MEANS AND *t*-TEST SCORES FOR SUCCESSFUL *vs.* UNSUCCESSFUL GROUPS ON THE "SMOKING QUESTIONNAIRE"

Scale	Group	*N*	Successful group $\overline{X}$	Unsuccessful group $\overline{X}$	Critical ratio
D	M	14	19.13	20.67	.64
	F	17	20.57	23.50	1.34
	Comb.	31	19.84	22.09	1.29
Hy	M	14	22.00	23.67	.55
	F	17	24.70	27.10	.86
	Comb.	31	23.35	25.39	.99
Pd	M	14	23.00	23.83	.33
	F	17	24.29	23.80[a]	—.22[a]
	Comb.	31	23.65	23.82	.10
Ma	M	14	19.63	20.00	.14
	F	17	22.29	17.70[a]	—1.86[a]
	Comb.	31	20.96	18.85[a]	—1.15[a]
Si	M	14	25.13	25.43	.09
	F	17	27.00	36.44	3.07**
	Comb.	31	26.07	30.94	2.14*
Ne	M	14	6.88	7.83	.45
	F	17	8.00	10.50	1.30
	Comb.	31	7.44	9.17	1.21

* Significant at .05.
** Significant at .005.
[a] Opposite to predicted direction.

TABLE 3

SCORES FOR SUCCESSFUL AND UNSUCCESSFUL WITHDRAWAL FOR FEMALES ON THE *Si* SCALE

Category	Sex	Cutting score	% correct	% incorrect	*N*
Successful	F	33 or lower	100.0	.0	7
Unsuccessful	F	34 or higher	78.0	22.0	10
Total			87.5	12.5	17

nificant difference between successful and unsuccessful withdrawers was found on the *Si* scale with female *S*s. Success of male smokers could not be predicted along any of the scale dimensions. The significance of the *Si* scale as a predictor of withdrawal success for females relates to earlier research (3) in which the females in a similar smoking-withdrawal group did not respond as well to educational therapy as did males. It may be that the characteristics ascribed to social introversion, such as shyness, modesty, and self-effacement, reduce the potential of a group setting for inducing successful withdrawal with females (8).

D. SUMMARY

An MMPI smoking questionnaire was derived and administered to persons volunteering for a smoking-withdrawal program. After three therapeutic

sessions, comparisons were made between the MMPI scores for persons who successfully quit smoking and those who could not terminate smoking. Although the unsuccessful *S*s scored higher than the successful *S*s on the *D, Hy, Pd, Si,* and *Ne* scales, only the *Si* scale with female *S*s was significantly different, with the unsuccessful females scoring higher than the successful females.

REFERENCES

1. LILIENFELD, A. M. Emotional and other selected characteristics of cigarette smokers and non-smokers as related to epidemiological studies of lung cancer. *J. Nat. Cancer Inst.,* 1959, **22**, 259-282.
2. MATARAZZO, J. P., & SASLOW, G. Psychological and related characteristics of smokers and non-smokers. *Psychol. Bull.,* 1960, **57**, 493-513.
3. SCHAUBLE, P. G., WOODY, R. H., & RESNIKOFF, A. Educational therapy and withdrawal from smoking. *J. Clin. Psychol.,* 1967, **23**, 518-519.
4. SCHUBERT, D. S. P. Personality implications of cigarette smoking among college students. *J. Consult. Psychol.,* 1959, **23**, 376-384.
5. ———. Arousal seeking as a central factor in tobacco smoking among college students. *Internat. J. Soc. Psychiat.,* 1965, **11**, 221-225.
6. SELTZER, C. C. Occupation and smoking in college graduates. *J. Appl. Psychol.,* 1964, **48**, 1-6.
7. STRAITS, B., & SECHREST, L. Further support of some findings about the characteristics of smokers and non-smokers. *J. Consult. Psychol.,* 1963, **27**, 282.
8. WELSH, G. S., & DAHLSTROM, W. G. An MMPI Handbook: A Guide to Use in Clinical Practice and Research. Minneapolis, Minn.: Univ. Minnesota Press, 1960.
9. WINNE, J. F. A scale of neuroticism: An adaptation of the MMPI. *J. Clin. Psychol.,* 1951, **7**, 117-122.

College of Education
Erickson Hall
Michigan State University
East Lansing, Michigan 48823

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 121-127.

RELEVANT AND IRRELEVANT CUES IN SPATIAL REVERSAL LEARNING*[1]

Department of Psychology, The University of Calgary

ROBERT G. WEYANT

A. INTRODUCTION

In a recent review of certain aspects of animal discrimination learning, Mackintosh (2) commented upon the relative lack of spatial reversal studies utilizing irrelevant cues. In the same paper, Mackintosh distinguished between one- and two-stage discrimination theories and presented evidence pertinent to the inadequacy of the one-stage continuity theory. The purpose of the present study was to provide data concerning the role of relevant and irrelevant cues in spatial reversal learning.

The most extreme position of one-stage continuity theory would state that animals should learn equally about all of the stimuli impinging upon their receptors. Thus, in a spatial learning situation, an animal which, within a reversal, has a visual signal consistently associated with reinforcement would be expected to learn not only the spatial response (left turn or right turn) but also the response of approaching the appropriate stimulus. If the reversal situation is so designed that such an animal must reverse not only its position response but also its visual discrimination, one might well expect such a reversal to involve a greater number of errors than would a similar reversal made by an animal trained with the visual stimuli irrelevant so that only the spatial response had to be reversed. This expectation does not necessarily arise from the fact that, presumably, two habits have to be reversed in the first case as opposed to only one in the second, but is suggested by comparisons of error scores in spatial and nonspatial (visual) reversal studies which have appeared in the literature (4).

B. METHOD

1. *Subjects*

The *S*s were 18 experimentally naive albino rats of the Wistar strain. All were between 80 and 120 days of age at the beginning of the study.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on October 30, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported, in part, by National Research Council of Canada Grant No. APA-183, "An examination of some complex learning phenomena."

2. *Apparatus*

The apparatus was a simple T-maze with the stem 24 inches long × 5 inches wide × 11 inches deep and each arm 16 inches long × 5 inches wide × 11 inches deep. The stem was painted dark gray; both arms and the choice-point were white. A transparent door 10 inches from the base of the stem formed a start-box and a similar door in each arm of the T prevented retracing. At the end of each arm, farthest from the choice-point, a clear white signal arm was mounted 3 inches from the floor of the maze. The light in either arm could be turned on independently of the other or both could be lighted simultaneously. When the light in one arm was on and the other was off, readings taken at the choice-point indicated an illumination level of 3.5 fc for the lighted arm and .2 fc for the unlighted arm. These were highly discriminable to a human observer. Food cups were located on the floor of the maze directly below the signal lights in each arm. The only other illumination in the room came from a dim bulb hung 56 inches above the floor of the choice-point which gave a reading of .1 fc at the choice-point when both signal lights were turned off.

3. *Procedure*

The *S*s were divided, by means of a table of random numbers, into three groups of six *S*s each. All groups were given original training and 10 reversals of a spatial (left *vs.* right) response. In all cases, the position of reinforcement was reversed when *S* met a criterion of 12 successive correct responses. One-half of *S*s in each group began the study with a left turn correct, one-half with a right turn correct.

In one of the groups (Group C) a visual cue (light on for one-half of *S*s, light off for one-half) was consistently correlated with the position of reinforcement. For instance, if in original training the correct response was "left" and the left-hand arm had the light lighted while the right-hand arm had the light unlighted, then when *S* reached criterion and the position of the reinforcement was switched to the right-hand arm the signal light would also switch to the right-hand arm. Thus, for *S*s in this group, the same visual cue (either light on or light off) was associated with the position of reinforcement throughout the original learning and all 10 reversals.

In a second group (Group I) a visual cue (light on for one-half of *S*s, light off for one-half) was consistently correlated with reinforcement on an intra-reversal basis, but not between reversals. For instance, if in original training the correct response was "left" and the left-hand arm had the light lighted while the right-hand arm had the light unlighted, then when *S* reached criterion and the position of reinforcement was switched to the right-hand arm the signal light would *not* switch to the right-hand arm, but would remain lighted in the left-hand arm. Thus, for *S*s in this group, a visual cue (either light on or light off) was associated with the position of reinforcement within a reversal, but the same visual cue was not associated with the position of reinforcement from one reversal to the next.

In a third group (Group U) the position of the lighted arm was varied on a LRRLRLLR schedule so that the visual cue was never consistently associated with either reinforcement or nonreinforcement. Thus, for *S*s in this group, visual cue was uncorrelated with the position of reinforcement, and so was irrelevant.

All *S*s were housed in individual cages and had free access to water at all times except while in the maze. A feeding schedule was maintained throughout the experiment on which *S*s were given 2.5 hours of free access to food each day, ending 17 hours prior to the experimental session of the following day. This schedule was maintained for three weeks prior to the start of the study. At the regular feeding time on the day prior to the pretraining period, *S*s also received a few pellets of the type which were to be used as reinforcement.

All *S*s had a pretraining period of three days during which they were given four trials in the maze each day on a free-forced schedule to insure that two of the four trials would be to the positive side and two to the negative. The cue conditions in each case were those appropriate to *S*'s group. During the remainder of the study, *S*s were given 12 free trials per day in blocks of four trials. Time between trials within a block was approximately 45 seconds, time between blocks of trials was approximately 10 minutes. Water was available to *S*s, who were restrained in their home cages during between-block periods. Reinforcement in the maze for all *S*s consisted of four pellets of Lab Rat Food (4.0 mm × 3.3 mm × 45 mg) produced by the P. J. Noyes Company.

C. RESULTS

The results of this study are summarized graphically in Figure 1. This figure shows the mean number of errors as a function of reversals. Unless otherwise stated, all statistical tests referred to in this paper were Duncan Range Tests at the .05 level.

During original training, Groups I and C (which received the same treatment) did not differ significantly from each other, while Group U did make a significantly greater number of errors. On the first reversal, the only significant difference was the one between Groups U and C, while on all subsequent reversals Groups U and I did not differ significantly from each other, but both differed significantly from Group C (except on the fourth reversal where the difference between Groups I and C barely missed significance).

All groups show a significant decrease in errors between the first and tenth reversals (t for related measures, $p = .01$).

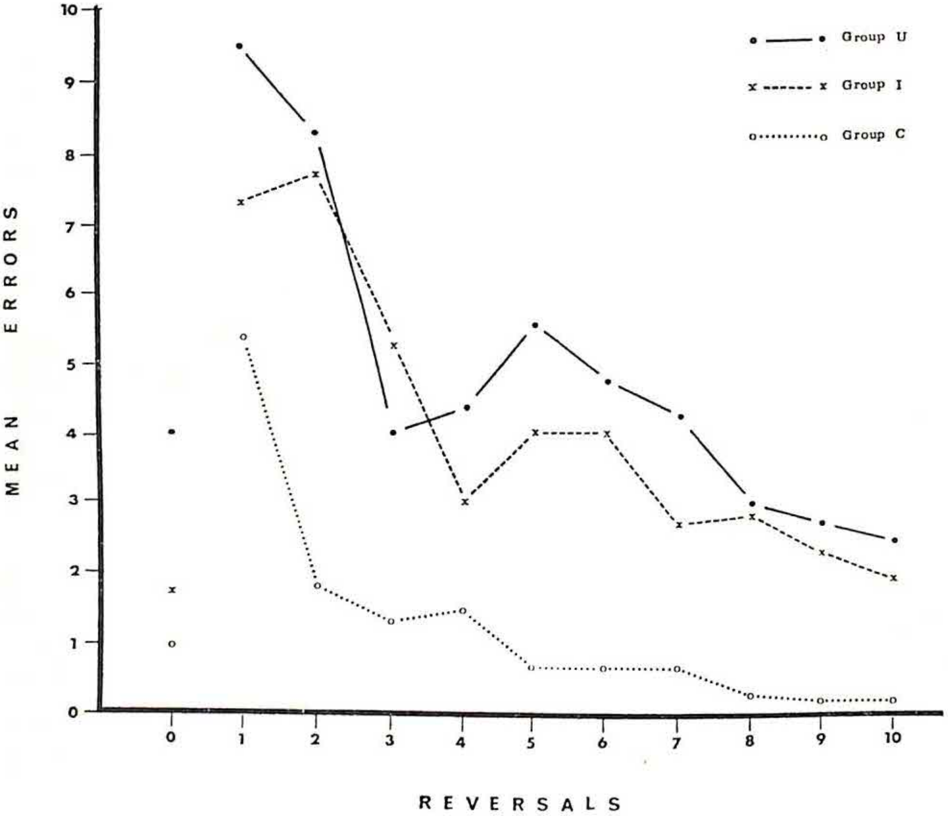


FIGURE 1
Mean Number of Errors as a Function of Reversals

D. Discussion

The results of this study do not confirm the prediction made from the point of view of a single-stage continuity theory. Group I, which presumably had to reverse both a position habit and a visual discrimination, did not make a greater number of errors than Group U, which only had to reverse the position habit. In fact, Group I shows *less* errors in nine out of 10 reversals (although not significantly less) than does Group U. This raises the question of whether the performance of Group U, which was presented with irrelevant signals, may have been disrupted by the presence of the uncorrelated visual cues. This question is crucial to any interpretation of, for instance, the original training results. The significant differences between Group U and Groups I and C during the original training could be due to either a disruptive effect of the uncorrelated cues on Group U or the ability of Groups I and C, even during original training, to utilize the correlated visual cues, or both factors.

Some light is shed on this question by unpublished data collected by the author in connection with another study run at a different time, but utilizing the same types of *S*s in the same apparatus. The only difference in procedure was that the lights in both arms were lighted all during original training and the 10 reversals. Thus, this group had no correlated visual cues, but also did not have any potentially distracting uncorrelated visual cues supplied by *E*. These data are presented graphically in Figure 2 along with the data from Group U. At no point do these groups differ significantly (*t*-test, .05 level).

It must, therefore, be concluded that under the conditions of this study there are no disruptive effects of irrelevant visual stimuli on the learning or reversal of a spatial discrimination. Thus, the fact that Groups I and C perform significantly better on original learning than does Group U, and the fact that Group C continues to perform significantly better over all 10 reversals, must be interpreted as due to the utilization of the visual cues by these groups in these circumstances.

The question still exists as to whether Group I learns to use its intracorrelated visual cues to decrease errors *within* a reversal, even though no detrimental effect appears from one reversal to the next. While this study does not provide an unambiguous answer to this question, there is some evidence that the answer might be affirmative. Dufort, Guttman, and Kimble (1) have provided evidence of one-error spatial reversals in rats. After a number of reversals, therefore, the major advantage of Group C over Group I would be expected to be on the first trial following a reversal when *S*s under Group C conditions can (and do) make errorless reversals. Comparisons were made of all three groups, eliminating from the data the first trial following each reversal. Interestingly enough, when this is done the differences between Groups I and C fail to attain significance on three of the last four reversals (Nos. 7, 9, and 10) although the differences between Groups U and C continue to attain significance (Duncan Range, $p = .05$). This could be interpreted—but does not conclusively demonstrate—that *S*s in Group I are making use of the visual cues on a within-reversal basis. The fact that this does not retard their ability to reverse the discrimination is interesting.

Finally, the failure of Group I to perform significantly better than Group U over the series of reversals, although Group I had clearly learned the relevant dimensions in the original learning (reversal O), calls into question the suggestion of Mackintosh and Mackintosh (3) that improvement over reversals is primarily a case of determining the correct discriminative dimension.

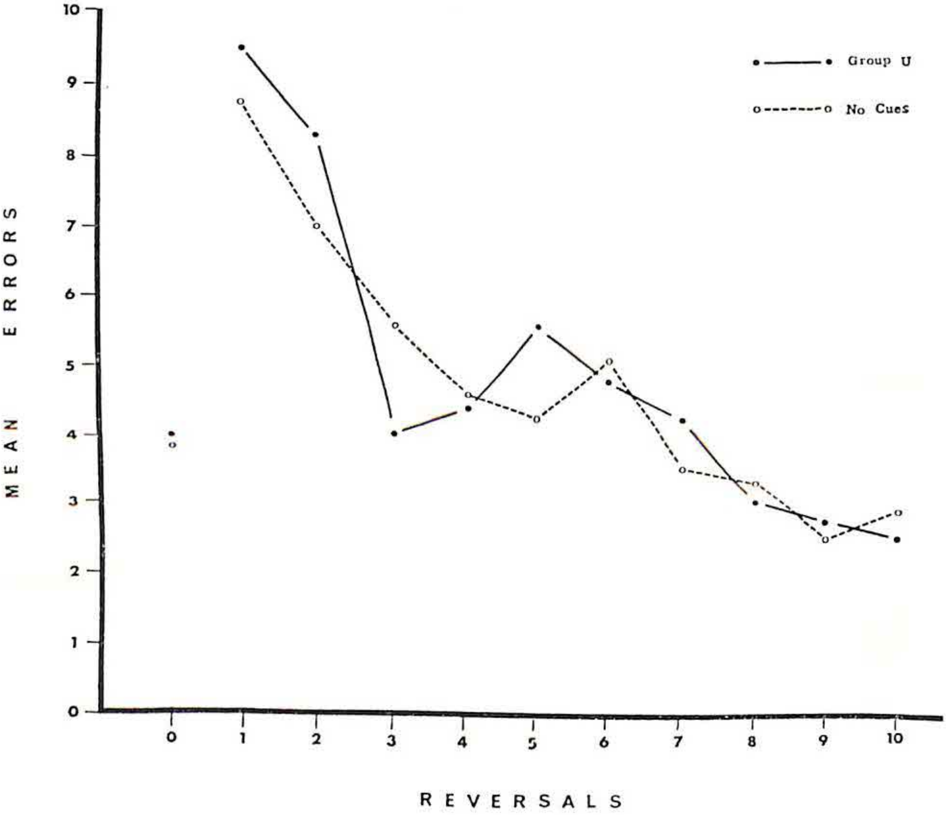


FIGURE 2
Mean Number of Errors as a Function of Reversals

E. Summary

Rats were given original training and 10 reversals in a spatial learning (T-maze) situation. Group C (correlated) had a visual cue consistently associated with the spatial location of the reinforcement. Group I (intracorrelated) had a visual cue associated with the spatial location of reinforcement within a reversal, but not consistently from one reversal to the next. Group U (uncorrelated) had visual cues presented in a pattern which was

not correlated with the spatial location of reinforcement. *S*s in Group C clearly made use of the correlated visual cues; there was some suggestion that *S*s in Group I might be using the visual cues on a within-reversal basis. No evidence was found for the disruptive effect of irrelevant visual stimuli on the learning or reversal of a spatial discrimination. An hypothesis derived from single-stage continuity theory was not supported.

REFERENCES

1. DUFORT, R. H., GUTTMAN, N., & KIMBLE, G. A. One-trial discrimination reversal in the white rat. *J. Comp. & Physiol. Psychol.*, 1954, **47**, 248-249.
2. MACKINTOSH, N. J. Selective attention in animal discrimination learning. *Psychol. Bull.*, 1965, **64**, 124-150.
3. MACKINTOSH, N. J., & MACKINTOSH, J. Performance of *Octopus* over a series of reversals of a simultaneous discrimination. *Anim. Behav.*, 1964, **12**, 321-324.
4. WEYANT, R. G. Reversal learning in rats as a function of the type of discrimination and the criterion of learning. *Anim. Behav.*, 1966, **14**, 480-484.

Department of Psychology
The University of Calgary
Calgary, Alberta, Canada

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 129-140.

A MATHEMATICAL REPRESENTATION OF TIME-VARIANT QUANTITATIVE BEHAVIOR*

Tucson, Arizona

JOHN VOEVODSKY

A. INTRODUCTION

Certain quantitative forms of repetitive behavior performed by biological systems of varying complexity are found to be accurately described by solutions obtained from two differential equations. These two equations express the relationship between the input and output of complex behavioral events as a function of time regardless of the complexity of the biological system. In addition to the general equations, two widely divergent specific cases are derived from the general behavioral equations. The differences between biological systems are obvious, as are the differences in the complexity of the individual behavioral events which they are capable of repeating. What is of interest is that, when the quantitative behavioral concepts of Stevens (12) are combined with the cybernetic concepts of Wiener (16), the biological systems which have been analyzed so far perform in a *similar* manner regardless of their level of organization or complexity of their behavioral events.

The literature reports special cases where specific mathematical formulae are shown to describe accurately the observed results. For example, in 1927 Hill (6) demonstrated the very small error between observed position as a function of time of an athlete and the calculated position of the athlete either running or broad jumping. Fitts (5) reports numerous other examples for humans performing analog tracking experiments. Studies performed on humans subject to suprathreshold pain indicate similar mathematical relationships (15). Since the specific formulae which describe these special cases plus the two cases presented here can all be derived from two ordinary differential linear equations, these equations are now considered.

B. SCALES FOR MEASUREMENT

In general, Stevens shows that quantification of behavior is readily achieved by the assignment of numerals to distinguishable behavioral events in the repetitive behavior of biological systems as a function of time (12). For

---

* Accepted by Ernest R. Hilgard of the Editorial Board, received in the Editorial Office, Provincetown, Massachusetts, on November 6, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

example, the total number of pecks up to any instant which a pigeon performs while undergoing experiments in a Skinner box (4) are distinguishable, repetitive, time-variant behavioral events (see Figure 1). How far a man runs (or the number of steps he has taken) measured in feet as a function of time are distinguishable events. A spoken number indicating that the pain has changed a uniform amount is a distinguishable event. Additionally, subevents which occur at a fixed ratio at some point during the performance of the principal behavioral event may also be used, such as the total number of seeds given to a pigeon for a given number of pecks, or the number of defense dollars (11) spent for building an army instead of the number of men in the army (see Figure 2). These behavioral events become the system variables

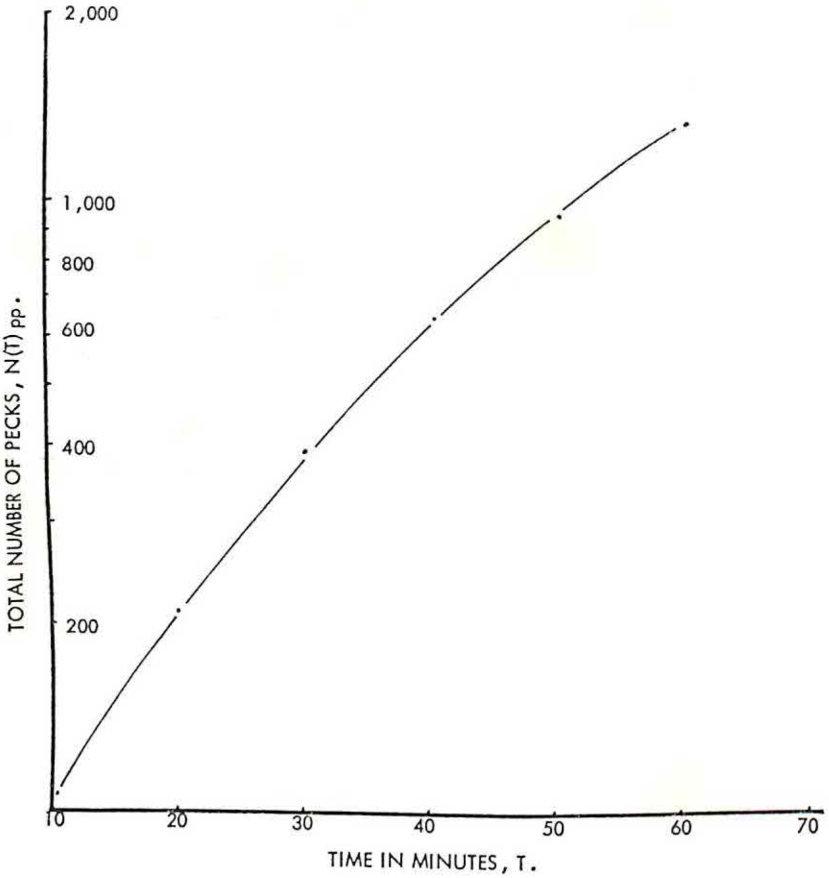


FIGURE 1
Total Number of Pecks as a Function of Time (4)

which are time variant. A biological system, then, is a combination of components (or units) where each unit has separate system variables that act together to form the overall system. It appears that the level of complexity of the biological organization depends upon the number of components that are acting together. The system can be analyzed either by considering the separate system variables for each component, or by just considering the system variables for the overall biological system (2). Which method is preferred apparently depends solely upon the information which the investigator wants from the analysis and the time he wishes to spend.

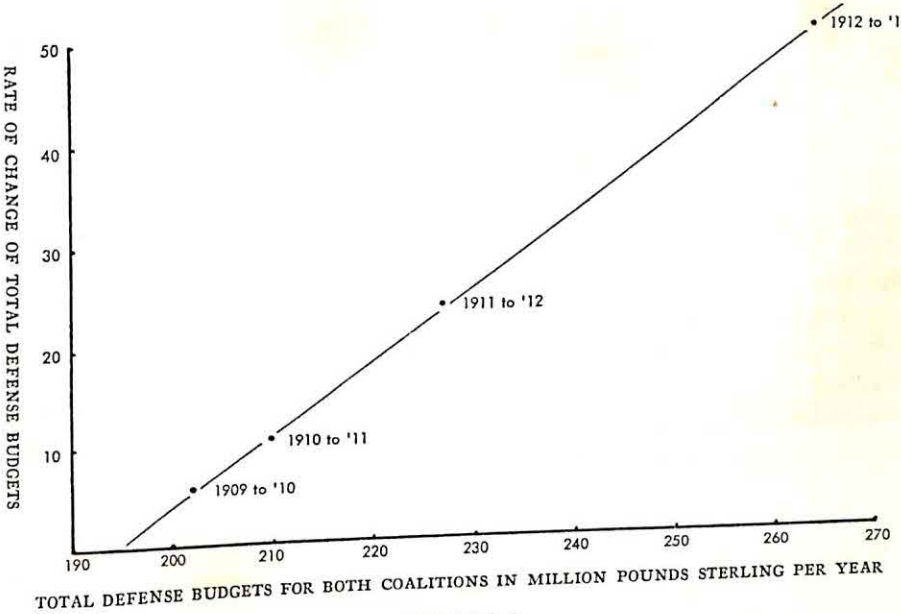


FIGURE 2
RATE OF CHANGE OF THE TOTAL DEFENSE BUDGETS FOR BOTH COALITIONS OF NATIONS IN MILLION POUNDS STERLING PER YEAR AS A FUNCTION OF THE TOTAL DEFENSE BUDGETS FOR BOTH COALITIONS OF NATIONS IN MILLION POUNDS STERLING PER YEAR (11)

C. IDEAL BIOLOGICAL SYSTEM

Simplification of any actual system is a necessary process in understanding the properties of the system. The resulting mathematical approximation to the real system is often referred to as the ideal system. The experiments already mentioned in this paper indicate that these biological systems may be ideally approximated initially by a linear second-order feedback control-systems equation of the type first conceived by Wiener (16), which has the general mathematical form:

$$F(t)_k = \sum_{j=1}^{m} \left[a_{jk} \frac{d^2R(t)_{jk}}{dt^2} + b_{jk} \frac{dR(t)_{jk}}{dt} + c_{jk}R(t)_{jk} \right], \qquad [1]$$

where a_{jk}, b_{jk}, and c_{jk} are constants and are assumed invariant when the system is not adapting; $R(t)_{jk}$ is the dependent variable of the k^{th} biological system and the j^{th} behavioral event which the system is in the process of repeating, and $F(t)$ is the forcing function, the precise meaning of which is considered in detail in the next section of this paper (but in short it can be considered as that stimulus which accelerates—positively or negatively—the repetition of certain behavioral events).

Thus $R(t)_{jk}$ is the total number of behavioral events performed by the biological system up to any given instant, t, from some arbitrary initial condition. For example, $R(t)_{jk}$ may be the total number of pecks from the beginning of the experiment up to any instant performed by a pigeon in a Skinner box, or it may be the total number of defense monies spent by the k^{th} coalition of nations as a function of time up to a given instant. If there are m number of subbehavioral events which when considered acting together make up the principal behavioral event, then these are summed—providing there is a fixed ratio between repetitive subevents. The failure of the biological system to complete any one of the subevents must result in failure of the organism to perform the total or principal behavioral event.

D. FORCING FUNCTION

Communication of the environment to the biological system occurs through its receptor cells or receptor units and its transmission system. Fortunately, there are many parallels to human perception in the magnitude of electrophysiological data from lower animals which are not now available in human data that permit a mathematical formulation of the perceptual consequences of some aspects of stimulus strength. (The strength of the stimulus as a function of time is by definition the forcing function.) For example, Ratliff (10) has found that the instantaneous frequency of discharge of the eye of *Limulus* is a function of the intensity of illumination for steady-state conditions and the rate of change of intensity during dynamic conditions. In fact, similarity in form of both dynamic and magnitude relationships appears to be a general characteristic of nerve cells (14) where the similarity has been found for kinesthetic cells (1, 9), muscle spindle (8), and motoneurons (7). All cells appear to respond to change or rate of change. This general characteristic may be summarized mathematically as follows:

$$F(t)_k = \pm 1 \sum_{i=1}^{n} \Big[{}^{a}_{ik}(S(t)_{ik} - S_{iok})^{\delta_{si}} + \beta_{ik}\left(\frac{dS(t)_{ik}}{dt} - \frac{dS_{iok}}{dt}\right)^{\delta_{di}} \Big] \qquad [2]$$

where the *algebraic* summation occurs over all sensory modalities. α_{ik} and β_{ik} are constants which depend on the past history of the k^{th} biological organism, and they vary when the organism adapts to the environment. $S(t)_{ik}$ is the stimulus at any instant impinging on the i^{th} sensory modality and $\dfrac{dS(t)_{ik}}{dt}$ is the rate of change of the stimulus, $S(t)_{ik}$. S_{iok} and $dS(t)_{iok}/dt$ are threshold values which the stimulus must exceed and below which $F(t)$ is zero. The constants, δ_{si} and δ_{di}, describe the characteristic negatively accelerating output magnitude of nerve cells to input magnitude, $S(t)_{ik}$ and $dS(t)_{ik}/dt$, which Stevens and Galanter found closely matched the data (13). In order to obtain a feeling for the algebraic summation over all sensory modalities a simple metaphor is sometimes useful: "loving corn and hating lima beans, we simply cannot make up our minds about succotash." In other words, Equation 2 states that the algebraic sum of that stimuli which is, for a particular biological system, unpleasant when subtracted from that other stimuli which is pleasant results in a difference, $F(t)$, which if other than zero will accelerate the repetition of the behavioral event in the manner described by Equation 1. Since both unpleasant and pleasant stimuli will positively accelerate behavior, a +1 or —1 must be placed at the beginning to take care of signs.

Quantification of the stimulus, $S(t)_{ik}$, proceeds in the same manner as the quantification of the behavioral events because in all cases part of the stimulus is composed of the behavioral event at that instant. Thus, for example, the degree of hunger of the pigeon is a stimulus and can be assumed a constant, C, if the seeds are small; the total number of seeds, $N(t)_{sp}$, the pigeon has received up to a given instant may be considered a stimulus; or the total defense expenditures, $N(t)_{ek}$, of the opposing coalition of nations—all these can serve as stimuli.

Since few biological systems have been analyzed by the method described in this paper, no hard and fast rules for determining which stimuli are important to a particular organism exist. One must proceed to the data and work from the data in order to obtain the particular forcing function which is causing the particular set of observed behavioral events. It is, of course, hoped that when more familiarity is obtained with this type of analysis, rules determining which stimuli are important to a particular organism will become known. Thus the parameters for Equation 2 have not been quantitatively specified. It is the form of the function that is inferred; the presumption is that the quantitative parameter must be determined empirically for a particular application of theory to practice.

E. Case I

For the first application of theory to practice, the simplest biological system (for which data readily exist) performing the least complex behavioral event repetitively as a function of time is chosen—a trained pigeon pecking in a Skinner box. For the second case, the most complex behavioral organism (coalitions of nations) performing for the most part the most complex behavioral event repetitively (arming for war) as a function of time is chosen. This choice is made on purpose in order to demonstrate the *similarity* in performance of biological systems which have obvious extreme differences in complexity of behavior as well as extreme differences in complexity of organization. The data for arming nations for World War I are also readily available. Furthermore, less extreme cases which have already been worked out in the literature are referenced (5, 6, 15).

A pigeon which has been deprived of food for some hours is placed in a Skinner box and is taught to peck a lever arrangement in order to receive food reward. If a "fixed ratio" schedule is used, then after a given number of pecks, $N(t)_{pp}$, the pigeon receives a seed. (The first subscript, p, stands for the event which is a peck, and the second subscript stands for the organism which is in this particular case a pigeon.) The fixed ratio is given as $N(t)_{pp} = u_1 N(t)_{sp}$. The experiment provides that the fixed ratio $u_1 = 50$ where the pigeon pecked 50 times for each seed. The seeds are so small that the bird's hunger is not decreased during the course of the experiment. The experiment, furthermore, is short enough so that the pigeon does not become fatigued. Finally, the purpose of the experiment is to determine after the bird begins to perform this particular series of behavioral events whether or not it will increase or decrease its pecking rate at any instant. Figure 1 shows the relationship that was actually obtained between $N(t)_{pp}$ and t.

Initially, the data show that at $t = 0$ the number of pecks, $N_{pp} = 0$, and also the slope, dN_{pp}/dt, can be assumed zero. As time progresses the bird's pecking rate increases until after 80 minutes it achieves a constant value of 60 pecks per minute, which it maintains until the end of the experiment.

The pigeon has two subbehavioral events, $m = 2$. The first is the eating of $N(t)_{sp}$ seeds so that $R(t)_{1p} = N(t)_{sp}$. The second subevent is pecking the lever a total number of pecks given by $N(t)_{pp}$ so that $R(t)_{2p} = N(t)_{pp}$. Equation 2 can then be written out in full as follows:

$$F(t)_p = \frac{(a_{1p} + u_1 a_{2p})}{u_1}\frac{d^2N(t)_{pp}}{dt^2} + \frac{(b_{1p} + u_1 b_{2p})}{u_1}\frac{dN(t)_{pp}}{dt} \qquad [3]$$

where in order to obtain the mathematical description of the observed data, c_{1p} must be set equal to zero, which can be interpreted as corresponding to the fact that the pigeon does not vary his response due to the total number of times that he has pecked. Equation 3 is the ideal systems equation for the pigeon in this particular experimental configuration.

In order to obtain the ideal forcing function, three sensory modalities must be considered, $n = 3$. The first modality are the sensors associated with the incoming seed reward:

$$\frac{dS(t)_{1p}}{dt} - \frac{dS_{1op}}{dt} = \frac{dN(t)_{sp}}{dt}$$

where $\alpha_{1p} = 0$ because the total number of seeds $N(t)_{sp}$ does not affect the observed behavior of the pigeon (the seeds are too small to be filling) and $\beta_{1p} > 0$ because the seeds are assumed pleasant.

The second modality is the feedback from the pigeon's total number of responses (pecks), $N(t)_{pp}$:

$$\frac{dS(t)_{2p}}{dt} - \frac{dS_{2op}}{dt} = \frac{dN(t)_{pp}}{dt}$$

where $\beta_{2p} > 0$ because the pecking rate does effect performance and is assumed to be pleasant, whereas $\alpha_{2p} = 0$ because the total number again is not fed back in the form of fatigue or other stimulus clues.

The third modality is the pigeon's hunger:

$$S(t)_{3p} - S_{3op} = C_p$$

where $\alpha_{3p} > 0$ and, since the seeds are small, the bird's hunger does not change, thereby allowing β_{3p} to be assumed zero.

Thus the stimulus strength is the algebraic sum of the input to the pigeon from the three sensory modalities; thereby Equation 2 can be written as:

$$F(t)_p = -1\left[\frac{(\beta_{1p} + u_1\beta_{2p})}{u_1}\frac{dN(t)_{pp}}{dt} - \alpha_{3p}C_p\right], \qquad [4]$$

where hunger has been assumed to be unpleasant and the pecking for reward is assumed pleasant. The difference between these system variables is used as the means of controlling the repetition rate of pecking. Also the fewer seeds the bird receives per peck the faster the pecking proceeds which is observed (4). The negative sign is used so that the unpleasant hunger (defined as negative) will give an accelerated response rate.

Equations 3 and 4 when equated give a function which describes a family of curves. The particular curve is obtained by considering the initial condition and final conditions thereby:

$$N(t)_p = 3324(e^{-1.81 \times 10^{-2t}} - 1) + 60t. \qquad [5]$$

The closeness of the fit of this equation to the experimental observations can be seen in Figure 1.

F. Case II

For the five years 1909-1913 prior to the outbreak of World War I, an armament race between two coalitions of nations occurred. France was allied with Russia and Germany was allied with Austria-Hungary. The two opposing alliances were roughly equal in size and at that time neither Italy nor Britain was in a definite alliance with either party.

The France-Russia coalition is designated by the subscript f on all variables, and the Germany-Austria-Hungary coalition by the subscript g. The total number of monetary units over a period of years spent on defense by France-Russia is $N(t)_{ef}$ where the subscript j has been changed to e in order that $N(t)_{ef}$ may be classified as monetary units (economics) spent on defense. The defense budget is given as $dN(t)_{ef}/dt$ for France-Russia coalition and $dN(t)_{eg}/dt$ for Germany-Austria-Hungary. The systems equation for the French coalition can then be written as

$$F(t)_f = a_{1f}\frac{d^2N(t)_{ef}}{dt^2} + b_{1f}\frac{dN(t)_{ef}}{dt} \qquad [6]$$

and for German coalition as

$$F(t)_g = a_{1g}\frac{d^2N(t)_{eg}}{dt^2} + b_{1g}\frac{dN(t)_{eg}}{dt} \qquad [7]$$

where again for some undefined reason the constants c_{1f} and c_{1g} are zero—in other words, the change in the defense budget does not apparently depend on how much has already been spent by either coalition.

In order to obtain the forcing function again three sensory modalities were considered, $n = 3$. The French coalition first: The first modality is the perception of the French coalition's own present defense budget:

$$\frac{dS(t)_{1f}}{dt} - \frac{dS_{1of}}{dt} = \frac{dN(t)_{ef}}{dt}$$

where $\alpha_{1f} = 0$ and $\beta_{1f} > 0$ thereby implying that only the present budget is considered and not the total investment.

The second sensory modality considered was the perception of the opposition's present defense budget:

$$\frac{dS(t)_{ef}}{dt} - \frac{dS_{2of}}{dt} = \frac{dN(t)_{eg}}{dt}$$

where again the opposition's total investment $N(t)_{eg}$ in defense seems to be ignored, thereby setting $\alpha_{2f} = 0$ and $\beta_{2f} > 0$.

The third sensory input was a constant input which might be considered a "drive" or an "ambition" term and seems to correspond to the hunger term in the previous example:

$$S(t)_{3f} - S_{3of} = C_f$$

where $\alpha_{3f} > 0$ and $\beta_{3f} = 0$ implies that for the extent of the race this drive function did not change appreciably.

The strength of stimulation to the French coalition is then given by substituting into Equation 2, thereby obtaining:

$$F(t)_f = -1 \left[\beta_{1f} \frac{dN(t)_{ef}}{dt} - \beta_{2f} \frac{dN(t)_{eg}}{dt} - \alpha_{3f} C_f \right] \qquad [8]$$

where for simplicity Stevens' power functions have been set equal to 1. Equation 8 may be interpreted as assuming that the French coalition takes comfort in the size of its defense budget but finds unpleasant the size of the German defense budget as well as its own "ambition," C_f. The negative sign is used in front of the brackets in order to give a positive acceleration caused by negative (unpleasant) forces.

Since both coalitions are about equal in size for the historical period under consideration, they apparently can be assumed equal in most every respect; and since the identical arguments hold for the German coalition, the name or label by which the coalition identifies itself can be changed from f to g and g to f in the above equations in order to obtain the system equation and stimulus strength equation for the German coalition. Thus by setting $c_{1f} = c_{1g} =$

0; $a_{1f} = a_{1g}$, $b_{1f} = b_{1g}$; $\beta_{1f} = \beta_{1g}$; $\beta_{2f} = \beta_{2g}$; $D_f = a_{3f}C_f/a_{1f}$; $D_g = a_{3g}C_g/a_{1g}$; and solving, the Richardson equation

$$\frac{d}{dt}\left(\frac{dN(t)_{cf}}{dt} + \frac{dN(t)_{cg}}{dt}\right) = \frac{b_{1f} + \beta_{1f} - \beta_{2f}}{a_{1f}}\left(\frac{dN(t)_{cf}}{dt} + \frac{dN(t)_{cg}}{dt}\right) + (D_f + D_g) \qquad [9]$$

is obtained and plotted with the data in Figure 2 where

$$\frac{a_{1f}(D_f + D_g)}{b_{1f} + 1_f - 2_f} = 194.$$

Equation 9 states that the rate of change of the total defense budgets of both coalitions is linearly proportional to the total defense budget of both coalitions during the period of time for which the above simplifying assumptions are historically valid. From 1909 to 1913 Equation 9 can be seen to be in close agreement with the actual data for that period.

In essence Figure 1 and Figure 2 show close agreement between theory and measurement. Therefore the entire purpose of the paper is to show that if the biological system is treated as a feedback control system as originally proposed by Weiner, which tends to maintain a prescribed relationship of one of its system variables to another, by comparing functions of these variables and using the *difference* as a means of controlling the repetition rate of its behavioral events as defined by Stevens, then close agreement between theory and measurement results.

G. CONCLUDING REMARKS

This paper has not only presented a set of differential equations, but also a means of applying them to biological systems in order that further understanding of the mechanisms of biological systems may eventually be realized. Furthermore, the empirical evidence for the derivation of these equations and for their existence has also been included.

Admittedly few analyses by the methods described in this paper have been performed. One must proceed to the data and work from there in order to begin to realize meaningful mathematical descriptions of time-variant repetitive behavioral events. No other methods are known to exist for analyzing such behavioral processes that have been described herein. Furthermore, the nature of the data keeps suggesting that analysis of this type may prove meaningful. For example, Figure 3 (3) shows the total number of deaths incurred by U.S. military personnel in connection with the conflict in Viet Nam resulting from action by hostile forces as a function of time. Worthy of special notice is how close the data points fall on the straight line drawn on the semilog axes. That the total number of deaths are distinguishable behavioral events, and that the mathematical function describing the events is easily written from the material presented in this paper, suggest that systems analysis of interacting nations might prove of value and interest to the nations involved with such current quantitative behavioral problems.

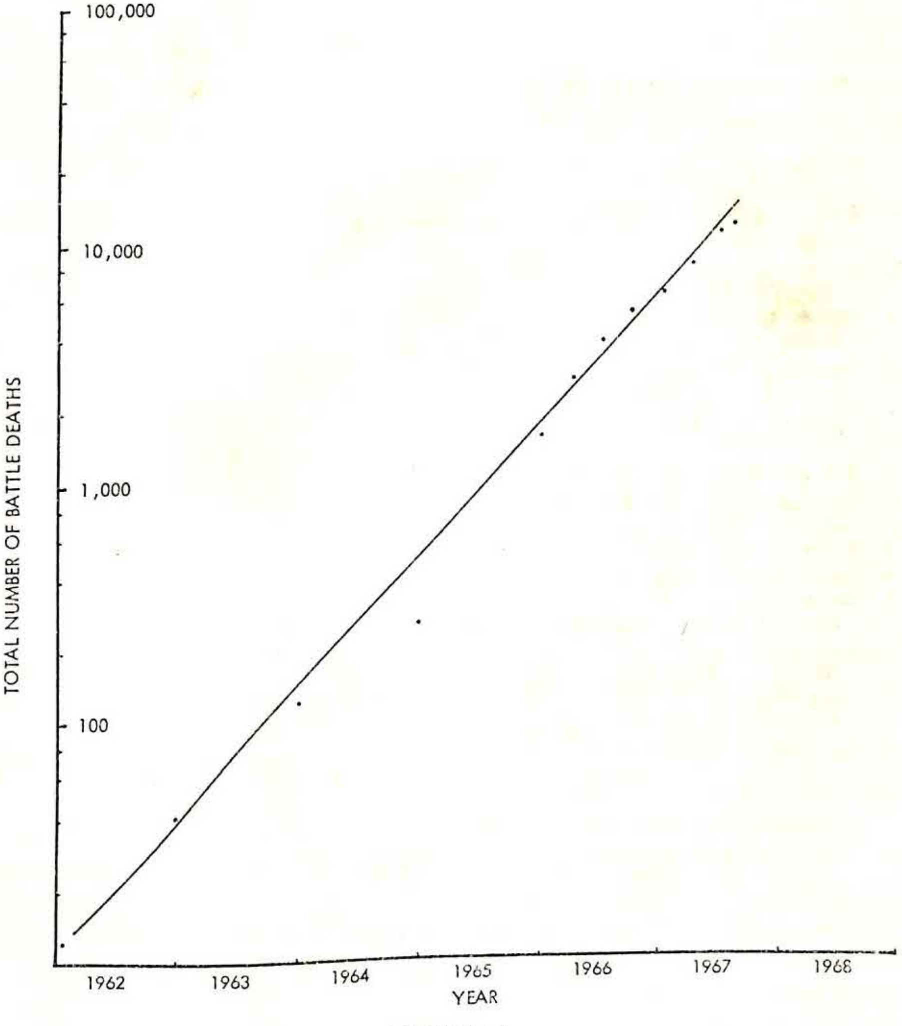


FIGURE 3
TOTAL NUMBER OF DEATHS INCURRED BY U.S. MILITARY PERSONNEL IN CONNECTION WITH THE CONFLICT IN VIET NAM RESULTING FROM ACTIONS BY HOSTILE FORCES AS A FUNCTION OF TIME (3)

H. SUMMARY

Certain quantitative forms of repetitive behavior performed by biological systems of varying complexity are found to be accurately described by solutions obtained from two differential equations. These two equations express the relationship between the input and output of complex behavioral events as a function of time regardless of the complexity of the biological system. What is of interest is that when the quantitative behavioral concepts of Stevens are combined with the cybernetic concepts of Wiener, the biological systems which have been analyzed so far perform in a similar manner regardless of their level of organization or complexity of their behavioral events.

REFERENCES

1. BOYD, I. A., & ROBERTS, T. D. M. Proprioceptive discharge from stretch-receptors in the knee-joint of the cat. *J. Physiol.*, 1953, **122**, 38.
2. D'AZZO, J. J., & HOUPIS, C. H. Feedback Control Systems Analysis and Synthesis. New York: McGraw-Hill, 1960. Pp. 1-129.
3. DIRECTORATE FOR STATISTICAL SERVICES. Office of the Secretary of Defense, Washington, D.C., August, 1967.
4. FERSTER, C. B., & SKINNER, F. B. Schedules of Reinforcement. New York: Appleton-Century-Croft, 1957. P. 47.
5. FITTS, P. M. Handbook of Experimental Psychology (ed. by S. S. Stevens). New York: Wiley, 1966. Pp. 1316-1323.
6. HILL, A. V. Are athletes machines? *Sci. Amer.*, 1927, **137**(2), 124-126.
7. HODGKIN, A. L. The local electric changes associated with repetitive action in a non-medullated axon. *J. Physiol.*, 1948, **107**, 165-181.
8. KATZ, B. Depolarization of sensory terminals and the initiation of impulses in the muscle spindle. *J. Physiol.*, 1950, **111**, 265-270.
9. MOUNTCASTLE, V. R. Handbook of Physiology, Section 1: Neurophysiology (Vol. 1, ed. by J. Field). Washington, D.C.: Amer. Physiol. Society, 1959. Pp. 387-427.
10. RATLIFF, F. Sensory Communication (ed. by W. A. Rosenblith). Cambridge, Mass.: M.I.T. Press, 1962. P. 196.
11. RICHARDSON, L. F. Arms and Insecurity. Pittsburgh, Pa.: Boxwood Press, 1960. Pp. 1-36.
12. STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley, 1966. Pp. 1-49.
13. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. Exper. Psychol.*, 1957, **54**, 377-411.
14. VOEVODSKY, J. A mathematical representation of interneuron behavior. Unpublished Doctoral dissertation, Stanford University, Stanford, California, 1965.
15. VOEVODSKY, J., COOPER, L. M., MORGAN, A. H., & HILGARD, E. R. The measurement of suprathreshold pain. *Amer. J. Psychol.*, 1967, **80**, 124-128.
16. WIENER, N. Cybernetics. Cambridge, Mass.: M.I.T. Press, 1962. Pp. 95-115.

826 West Cresta Loma
Tucson, Arizona 85704

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 141-143.

AUTHORITARIANISM, LOCUS OF CONTROL, AND RISK TAKING*[1]

Department of Psychology, University of Iowa

Robert A. Baron

A. Introduction

The authors of *The Authoritarian Personality* (1) attribute certain characteristics to High F individuals which may be related to conservative behavior in risk-taking situations. Among these are conventionalism, superstition, intolerance of ambiguity, and feelings of threat and insecurity. Indirect evidence in support of the hypothesis that high scores on the F scale are related to conservative behavior in risk-taking situations has been obtained in several experiments (2, 5, 7). More direct support for the above hypothesis has been obtained in a recent study by Raynor and Smith (4). Thus, one purpose of the present experiment was to examine the relationship between scores on the F scale and behavior in risk-taking situations. It was hypothesized that a positive correlation would exist between scores on the F scale and scores on a measure of risk-taking indicative of conservatism in a risk-taking situation.

Rotter, Seeman, and Liverant (6) have developed a scale to measure the extent to which an individual attributes control of reinforcement to his own behavior or to external forces. The higher an individual's score on this scale, the greater is his tendency to attribute control of his reinforcements to luck or chance. Previous evidence on the relationship between scores on this scale (the I-E scale) and behavior in risk-taking situations has been contradictory (3, 8). Thus, a second purpose of the present experiment was to examine the relationship between scores on the I-E scale and behavior in risk-taking situations. Because the I-E scale attempts to measure the individual's belief in luck or chance, it was hypothesized that a positive correlation would exist between scores on this scale and scores on a measure of risk-taking indicative of conservatism in a risk-taking situation.

The measure of risk taking chosen to test the above hypotheses was the choice dilemma problems devised by Wallach and Kogan (9). In these prob-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 8, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to express his appreciation to Drs. P. Murdoch and S. Arenson for their assistance in the preparation of this manuscript.

tems the individual indicates the *minimum probability* of success he would require before recommending certain risky actions.

B. METHOD

Seventy-five undergraduate students enrolled in sections of general psychology at the University of Iowa (58 females, 17 males) completed a true-false form of the F scale, the I-E scale, and the 12 Kogan and Wallach choice dilemma problems. These response measures were administered in counterbalanced order. Scores for the problems were the sum of the probabilities *S*s endorsed on all 12 items. The higher this sum, the more conservative were *S*s' recommendations judged to be. Pearson product-moment correlations between the three response measures were calculated with the aid of an IBM 7044 computer.

C. RESULTS

The correlation between scores on the F scale and the Kogan and Wallach problems was significant ($r = +.23$, $p < .05$), as was that between the scores on the I-E scale and these problems ($r = +.37$, $p < .01$). Thus, the higher an individual's score on either of the two personality measures, the higher his score on the problems tended to be. The correlation between scores on the F and I-E scales was not significant ($r = +.19$, $p > .10$). This finding indicates that scores on these two instruments are largely independent.

To examine further the relationship between scores on the F and I-E scales and scores on the Kogan and Wallach problems, mean scores on the problems for the 10 *S*s scoring highest and the 10 *S*s scoring lowest on these scales were compared by means of *t*-tests. The mean scores for the *S*s scoring high and low on the F scale were 69.40 and 60.70, respectively. This difference was not significant ($t = 1.32$, $df = 18$, $p > .20$). The mean scores for *S*s scoring high and low on the I-E scale were 71.30 and 54.00, respectively. This difference was significant ($t = 3.02$, $df = 18$, $p < .01$). Thus, high scorers on the I-E scale scored significantly higher on the Kogan and Wallach problems than low scorers.

D. DISCUSSION

The hypotheses that positive correlations would exist between scores on the F and I-E scales and scores on a measure of risk-taking indicative of conservatism in a risk-taking situation were supported. Thus, authoritarianism and belief in external control of reinforcement appear to have some positive relationship with conservative behavior in risk-taking situations. In view of the low correlations obtained, however, it would seem that risk taking, as measured in this experiment, has only a minor dependence on the variables measured by the F and I-E scales.

The finding that high and low scorers on the I-E scale differed significantly in their responses to the problems, while high and low scorers on the F scale did not, indicates that scores on the I-E scale may be more closely related to behavior in the type of situation described by the Kogan and Wallach problems than scores on the F scale. Further experimentation is necessary to determine if similar results will be found in different risk-taking situations.

E. SUMMARY

An experiment was performed to test the hypotheses that positive correlations would exist between scores on the F and I-E scales and scores on a measure of risk-taking indicative of conservatism in a risk-taking situation. The measure of risk-taking used was the Kogan and Wallach choice dilemma problems. Although both hypotheses were confirmed, the correlations obtained were quite low in absolute terms. It may be concluded, therefore, that factors other than those measured by the F and I-E scales are of greater importance in predicting performance on the Kogan and Wallach problems.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. The Authoritarian Personality. New York: Harper, 1950.
2. KOGAN, N., & WALLACH, M. Risk Taking. New York: Holt, Rinehart, & Winston, 1964.
3. LIVERANT, S., & SCODEL, A. Internal and external control as determinants of decision making under conditions of risk. *Psychol. Rep.*, 1960, **7**, 59-67.
4. RAYNOR, J. O., & SMITH, C. P. Achievement related motives and risk-taking in games of skill and chance. *J. Personal.*, 1966, **34**, 176-198.
5. RIM, Y. Who are the risk-takers? *Person. Admin.*, 1966, **29**, 26-30.
6. ROTTER, J. B., SEEMAN, M., & LIVERANT, S. Internal versus external control of reinforcement: A major variable in behavior theory. In N. Washburne (Ed.), *Decisions, Values, and Groups* (Vol. 2). New York: Pergamon Press, 1962. Pp. 473-516.
7. SCODEL, A., RATOOSH, P., & MINAS, J. S. Some personality correlates of decision making under conditions of risk. *Behav. Sci.*, 1959, **4**, 19-27.
8. STRICKLAND, L. H., LEWICKI, R. J., & KATZ, A. M. Temporal orientations and perceived control as determinants of risk-taking. *J. Exper. & Soc. Psychol.*, 1966, **2**, 143-151.
9. WALLACH, M. A., & KOGAN, N. Sex differences and judgmental processes. *J. Personal.*, 1959, **27**, 555-567.

Department of Psychology
University of Iowa
Iowa City, Iowa 52240

lems the individual indicates the minimum probability of success he would require before recommending certain risky actions.

B. METHOD

Seventy-five undergraduate students enrolled in sections of general psychology at the University of Iowa (58 females, 17 males) completed a true-false form of the F scale, the I-E scale, and the 12 Kogan and Wallach choice dilemma problems. These response measures were administered in counterbalanced order. Scores for the problems were the sum of the probabilities *S*s endorsed on all 12 items. The higher this sum, the more conservative were *S*s' recommendations judged to be. Pearson product-moment correlations between the three response measures were calculated with the aid of an IBM 7044 computer.

C. RESULTS

The correlation between scores on the F scale and the Kogan and Wallach problems was significant ($r = +.23$, $p < .05$), as was that between the scores on the I-E scale and these problems ($r = +.37$, $p < .01$). Thus, the higher an individual's score on either of the two personality measures, the higher his score on the problems tended to be. The correlation between scores on the F and I-E scales was not significant ($r = +.19$, $p > .10$). This finding indicates that scores on these two instruments are largely independent.

To examine further the relationship between scores on the F and I-E scales and scores on the Kogan and Wallach problems, mean scores on the problems for the 10 *S*s scoring highest and the 10 *S*s scoring lowest on these scales were compared by means of *t*-tests. The mean scores for the *S*s scoring high and low on the F scale were 69.40 and 60.70, respectively. This difference was not significant ($t = 1.32$, $df = 18$, $p > .20$). The mean scores for *S*s scoring high and low on the I-E scale were 71.30 and 54.00, respectively. This difference was significant ($t = 3.02$, $df = 18$, $p < .01$). Thus, high scorers on the I-E scale scored significantly higher on the Kogan and Wallach problems than low scorers.

D. DISCUSSION

The hypotheses that positive correlations would exist between scores on the F and I-E scales and scores on a measure of risk-taking indicative of conservatism in a risk-taking situation were supported. Thus, authoritarianism and belief in external control of reinforcement appear to have some positive relationship with conservative behavior in risk-taking situations. In view of the low correlations obtained, however, it would seem that risk taking, as measured in this experiment, has only a minor dependence on the variables measured by the F and I-E scales.

The finding that high and low scorers on the I-E scale differed significantly in their responses to the problems, while high and low scorers on the F scale did not, indicates that scores on the I-E scale may be more closely related to behavior in the type of situation described by the Kogan and Wallach problems than scores on the F scale. Further experimentation is necessary to determine if similar results will be found in different risk-taking situations.

E. SUMMARY

An experiment was performed to test the hypotheses that positive correlations would exist between scores on the F and I-E scales and scores on a measure of risk-taking indicative of conservatism in a risk-taking situation. The measure of risk-taking used was the Kogan and Wallach choice dilemma problems. Although both hypotheses were confirmed, the correlations obtained were quite low in absolute terms. It may be concluded, therefore, that factors other than those measured by the F and I-E scales are of greater importance in predicting performance on the Kogan and Wallach problems.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. The Authoritarian Personality. New York: Harper, 1950.
2. KOGAN, N., & WALLACH, M. Risk Taking. New York: Holt, Rinehart, & Winston, 1964.
3. LIVERANT, S., & SCODEL, A. Internal and external control as determinants of decision making under conditions of risk. *Psychol. Rep.*, 1960, **7**, 59-67.
4. RAYNOR, J. O., & SMITH, C. P. Achievement related motives and risk-taking in games of skill and chance. *J. Personal.*, 1966, **34**, 176-198.
5. RIM, Y. Who are the risk-takers? *Person. Admin.*, 1966, **29**, 26-30.
6. ROTTER, J. B., SEEMAN, M., & LIVERANT, S. Internal versus external control of reinforcement: A major variable in behavior theory. In N. Washburne (Ed.), *Decisions, Values, and Groups* (Vol. 2). New York: Pergamon Press, 1962. Pp. 473-516.
7. SCODEL, A., RATOOSH, P., & MINAS, J. S. Some personality correlates of decision making under conditions of risk. *Behav. Sci.*, 1959, **4**, 19-27.
8. STRICKLAND, L. H., LEWICKI, R. J., & KATZ, A. M. Temporal orientations and perceived control as determinants of risk-taking. *J. Exper. & Soc. Psychol.*, 1966, **2**, 143-151.
9. WALLACH, M. A., & KOGAN, N. Sex differences and judgmental processes. *J. Personal.*, 1959, **27**, 555-567.

Department of Psychology
University of Iowa
Iowa City, Iowa 52240

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 145-149.

PSYCHOSEXUAL CONFLICT IN ADOLESCENT ENURETICS*

Morristown, New Jersey

JOSEPH DELUCA

A. INTRODUCTION

The experience of psychoanalytic workers has suggested the presence of three psychological factors operative in the psychodynamics of enuresis: the fantasy of cloacal coitus (5, 6, 8, 9, 10), the fear of castration (7, 10), and thoughts of pubic hair (7, 10). The following formulation of the psychodynamics of enuresis was offered by Palm:

> Psychologically enuresis is a substitute for the unconsciously desired oedipal sex act which the enuretic patient conceives of as a particular form of cloacal contact. Enuresis is, at the same time, a defense mechanism against such wish, the fulfillment of which is feared to result in castration. Enuresis, therefore is the symptomatic expression of a particular modification of the oedipal conflict. This conflict may be reactivated by the sight of pubic hair on the basis of certain unconscious associations . . . (10, p. 179).

The reports of Freud (5, 6), Gerard (7), Katan-Angel (9), and Herbert (8) were all based on their clinical cases, while Palm's findings were based on a symbolic analysis of Rorschach responses. The purpose of this study was to make a more comprehensive and less subjective assessment of the psychosexual dynamics of enuretics than has been previously done.

B. METHOD

1. *Measures*

The measuring instrument used was the Blacky Test, as factor analyzed by Blum (1). The Blacky Test yields scores on 30 factors that were delineated as a result of the factoring of the 11 dimensions of psychosexual conflict that the Blacky Test originally purported to measure. The dimensions and the factors associated with each are delineated in Table 1.

2. *Subjects*

All *S*s had been recently inducted into the United States Army and were undergoing basic training at Fort Jackson, South Carolina. There were no

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 8, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

TABLE 1
DIMENSIONS AND FACTORS WITHIN DIMENSIONS MEASURED BY THE BLACKY TEST

Dimension	Factor
I. Oral eroticism	a. oral craving b. oral rejection c. sugar coating
II. Oral sadism	a. playfulness b. supply seeking c. resentment re oral deprivation
III. Anal sadism	a. exploitation b. choosing neutral responses c. denial of anal preoccupation
IV. Oedipal intensity	a. undisguised oedipal involvement b. disguised oedipal involvement
V. Masturbation guilt	a. fear of punishment re masturbation b. concern over sexual maturation c. denial of masturbation guilt
VI. Castration anxiety	a. overwhelming castration conflict b. minimizing castration anxiety
VII. Identification process	a. father as preferred identity object b. mother as preferred identity object c. evasion of identity issue
VIII. Sibling rivalry	a. overt hostility toward sibling and mother b. reaction formation toward sibling rivalry c. rejection in favor of sibling
IX. Guilt feelings	a. partial denial of guilt b. guilt ridden hostility toward sibling c. qualification of pervasive guilt
X. Ego ideal	a. overtly positive perception b. negative perception of self and friend
XI. Love object	a. mother-surrogate as love object b. heterosexual fantasy c. narcissism

significant differences in age, education, or intelligence as measured by the Army Classification Battery between the two groups. The mean age, education, and equivalent WAIS-*IQ* score for the enuretic group was 18.0, 11.0, and 98, respectively. The controls' mean age, education, and intelligence were 18.4, 11.5, and 99, respectively.

All enuretic *S*s ($N = 20$) had a history of chronic bed-wetting since childhood and were referred to the Mental Hygiene Consultation Service of the United States Army Hospital by their unit commander between the first and fifth weeks of basic training.

Controls ($N = 40$) consisted of trainees, also between their first and fifth weeks of basic training, who were temporarily confined to the hospital for bad colds.

No *S* was used in the study, either enuretic or control, who had a history of organicity[1] or psychiatric hospitalization, who had the equivalent of a WAIS-*IQ* of 80 or below, or who displayed clinical evidence of psychosis.

3. *Procedure*

Each enuretic, after having met the criteria for selection, was seen by the examiner for an interview. The main purpose of the interview was to establish rapport. During the interview it was explained that he was going to be asked to do some things that would probably contribute to an understanding of bed-wetting. The purpose of the above explanation was to encourage each enuretic *S* to perceive the test experience not as one in which evidence was being sought to censure him but rather one in which evidence to stimulate understanding and acceptance was the main purpose. The same procedure was followed with the control *S*s except that they were not informed that the study was concerned with enuresis.[2] This was done to avoid having them feel that *E* might be looking for evidence of enuretic tendencies within them. Thus information regarding the nature of the research was kept from the control *S*s to maximize uninhibited responding to the test items. All *S*s were cooperative.

4. *Scoring*

The scoring of each Blacky protocol was carried out on two separate occasions: the first immediately after the administration and the second a few weeks later, with the identity of the *S* unknown. Those protocols on which there was a disparity in scoring (1 per cent of the cases) were scored a third time. The scoring method used was according to Blum (1).

C. RESULTS

A score falling in the upper half of the range of all possible scores for a particular factor was the criterion for disturbance on that factor, while a score falling in the lower half of the range was the criterion for nondisturbance. Thus a score of 4-7 inclusive would indicate disturbance on a factor where the possible range was 0-7.

---

[1] All CNS involvements (head injuries, brain surgery, etc.) were eliminated. There were no *S*s with known history of bladder disease or kidney dysfunction.

[2] Each control *S* was told that he was going to be asked to do some things for research purposes. No *S* in the control group challenged the *E* or questioned this statement.

Results revealed that there were 22 factors on which at least one enuretic *S* showed disturbance. There were 24 factors on which at least one control *S* showed disturbance. However, the enuretics, as a group, were not significantly different from the controls (χ^2, $df = 1$, $p < .05$) in degree of disturbance on any one factor.

D. DISCUSSION

The findings of the authors cited regarding the dynamics of enuresis do not appear to be supported as the enuretics were not significantly different from the controls on any one factor. There is no substantial evidence therefore that the group of enuretics studied were dealing with psychosexual conflicts that were any different in kind or intensity from those which the controls were dealing with.

Differences between this study and the others cited may be due to (*a*) differences in the *S*s that were used (not clearly defined in previous studies or inadequate *N:* e.g., single case), (*b*) differences in the validity of the techniques of observation used (interview and Rorschach symbols *vs.* the factored Blacky Test), or (*c*) to the possibility that the enuretics were, in fact, not different in terms of conflicts being dealt with from the controls. The last possibility is supported by some previous research (2, 3, 4).

E. SUMMARY

Previous investigators have implicated three psychological factors in the psychodynamics of enuresis: the fantasy of cloacal coitus, castration fear, and the sight of pubic hair. Most studies to date have been based on a few clinical cases. This study compared the performance of enuretics with controls on the 30 factors derived from the factor analyses of the eleven dimensions of psychosexual conflict measured by the Blacky Test. Results suggest that enuretics are not defending against psychosexual conflicts that are any different from those of controls. Implications for theory are considered.

REFERENCES

1. BLUM, G. S. A guide for research use of the Blacky pictures. *J. Proj. Tech.,* 1962, **26**, 1-29.
2. DELUCA, J. N. The dynamic structure of the sexual deviations. Paper read at the VIth International Congress on Rorschach, Paris, France, July, 1965.
3. ———. The structure of homosexuality. *J. Proj. Tech.,* 1966, **30**, 187-191.
4. ———. Enuresis in adolescence: An experimental study of personality structure. Paper read at the VIth International Congress on Child Psychiatry, Edinburgh, Scotland, July, 1966.
5. FREUD, S. The interpretation of dreams. In A. A. Brill (Ed.), *The Basic Writings of Sigmund Freud.* New York: Random House, 1938.

6. ———. On the transformation of instincts with special reference to anal erotism. In *Collected Papers.* London: Hogarth Press, 1938.
7. GERARD, M. Enuresis: A study in etiology. *Amer. J. Orthopsychiat.,* 1939, **9**, 25.
8. HERBERT, S. The psychogenetic root of enuresis. *Psychoanal. Rev.,* 1922, **9**, 34.
9. KATAN-ANGEL, H. From the analysis of a bedwetter. *Psychiat. Quart.,* 1935, **4**, 56.
10. PALM, R. The psychodynamics of enuresis. *Amer. Imago,* 1953, **10**, 167-179.

290 Madison Avenue
Morristown, New Jersey 07960

Published as a separate and in *The Journal of Psychology,* 1968, **68**, 151-157.

SYMPTOM PATTERNS IN PSYCHIATRIC PATIENTS: A GOAL-DIRECTED APPROACH TO PSYCHIATRIC SYMPTOMATOLOGY* [1]

Miami University, the University of Denver, and the University of Colorado

ALVIN MAHRER, PATRICIA STEWART, JOHN HORN, AND DENNIS LIND

A. INTRODUCTION

This study was designed to determine whether the psychological problems-symptoms-complaints of hospitalized psychiatric patients occur in identifiable patternings which, rather than fitting into the current psychiatric diagnostic system, represent major behavioral avenues toward achieving hospitalization and psychological goals within a psychiatric setting.

Nearly all studies which have in the past attempted to extract major symptom patterns of psychiatric patients have been based upon the standard psychiatric nosological approach and, accordingly, have been involved almost entirely with modifications and variations of traditional psychiatric dimensions (1, 2, 3, 4, 6, 9, 10, 11, 12, 16, 19, 21, 22, 23, 24). The present study represents an alternative to the traditional psychiatric approach, and, rather than dealing with symptoms of "disease," measures symptomatology as goal-directed behavior (5, 13, 14, 17, 18, 20). Within the goal-directed approach, the major patterns of behavioral symptomatology are seen as the primary avenues of gaining hospital admission and of moving toward the achievement of psychological goals within the hospital setting.

B. CRITIQUE OF TRADITIONAL PSYCHIATRIC APPROACH

The traditional approach begins with an explicit acceptance of the psychiatric diagnostic nomenclature. Within this given framework, the research goal has been to identify the component factors which comprise such traditional psychiatric entities as depression (6), hypochondriasis (16), schizophrenia (4, 23), and psychosis (3, 11, 19, 21, 22). The research instruments which have been used are composed of items extracted from psychiatric diagnostic cate-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 9, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Data for the present study were obtained from the Denver Veterans Administration Hospital while the senior author was on the staff of that institution. The authors wish to express their appreciation to Dr. Edward Simmel for his editorial assistance.

gories as reflected in standard psychiatric textbooks, manuals of psychiatric illness, and handbooks of mental disorders (1, 2, 4, 11, 16, 19, 21).

In contrast, a goal-directed approach rejects such psychiatric entities in preference for constructs extracted from the domain of behaviors related to motives, needs, and goals. The goal-directed approach seeks to develop research instruments which measure these motives, needs, and goals in terms of behavioral problems-symptoms-complaints directed toward hospital admission and toward psychological goals within the hospital setting. Thus the authors' approach searches for communalities and patternings from a different complementary vantage point in which (*a*) items are geared toward behavioral problems-symptoms-complaints, rather than items extracted from psychiatric disease-nosology; and (*b*) the aim is to extract and identify major behavioral patternings rather than to differentiate existing psychiatric diagnostic categories.

C. METHOD

1. *Subjects*

*S*s were 302 male psychiatric patients, hospitalized during the period from 1959-1965 at the Denver Veterans Administration Hospital. As far as was practicable, approximately every eighth admission was included in the sample regardless of psychiatric diagnosis, excluding readmissions and patients who were discharged within three days of admission.

2. *Procedure*

Each *S* was individually administered a checklist, the Psychological Problem Inventory (PPI), consisting of 150 items. The method of derivation (15) was designed to provide *S*s with a simple, straightforward means of defining the broad scope of their psychological problems from their own self-reported point of view, in concrete, behavioral terminology. The instructions allowed *S*s to indicate any number of items relating to their present clinical state.

From the 150 items of the checklist, a subsample of 84 items was selected for primary analysis. The basis of selection was item variance; items with small variance were eliminated to the point where the analyses could be run on the remainder.

Product moment correlations were calculated; principal axes factors were determined using unities in the main diagonal and adopting the "root-one" criterion for determining the number of factors (7). Six factors were indicated by this criterion. These were rotated to an approximation of simple structure using the Varimax (8) procedure. A summary of the rotated results is given in Table 1.

TABLE 1
FACTORS OF SELF-REPORTED PSYCHOLOGICAL PROBLEMS OF PSYCHIATRIC PATIENTS

Item	Loading
Factor I	
1. My arm muscles hurt me.	.60
2. When I walk, my hip pains me very much.	.58
3. Sometimes my ears feel like there is a roaring in them.	.50
4. At night when I should be sleeping I pace the floor.	.44
5. There are times when I can't see well, and things look hazy and unclear.	.43
6. When I am in a car, I sometimes have to think hard to remember where I am going.	.43
7. Sometimes my body seems to be warm all over.	.41
Factor II	
1. When I walk, I feel unsteady and I'm afraid I will fall down.	.59
2. I have a pain in the back of my neck.	.55
3. I have a pain in the muscles of my legs.	.55
4. After being on my feet awhile, my legs get weak and I have to rest.	.53
5. I have a pain in the upper part of my back.	.51
6. I get so scared that I'm afraid to drive a car.	.51
7. My whole body shakes and I can't stop it.	.50
8. I can't have an erection at times.	.44
Factor III	
1. I can't stop my drinking.	.61
2. Even though I don't want to, I spend my money on alcohol.	.49
3. When I get drunk, I throw up.	.49
4. I get up two or three times during the night.	.42
Factor IV	
1. When I'm with a group of people I watch the talkers but I can't talk.	.60
2. When another person talks to me I feel they really don't want to.	.58
3. I can't seem to hold a job for longer than six months.	.53
4. I can't hold a job; my boss fires me soon after I get a job.	.52
5. When I talk to women I get embarrassed and can't talk.	.50
6. I can't work as fast as some of the people I work with.	.49
7. When another person is critical of me my face feels tight and unnatural.	.48
8. Some people at work think I am odd.	.47
9. When I'm about to say something in a group, my heart pounds hard.	.45
10. Sometimes my tongue feels thick and tight.	.42
Factor V	
1. When I argue with my wife I complain that she doesn't understand me.	.69
2. I haven't had intercourse with my wife in the last month.	.59
3. I struck my wife in the face.	.57
4. Often my eyes have tears in them.	.50
5. I'm sure that I can't add a column of figures like I used to.	.43
6. Often I have to fight back tears in my eyes.	.43
7. When I get really angry I slam doors as hard as I can.	.42
Factor VI	
1. I feel like I want to hit some of the men I work with.	.61
2. Sometimes I feel like I might kill someone.	.61
3. I feel like strangling someone at times.	.59
4. I give people dirty looks.	.53
5. My wife is afraid I will hit her.	.50
6. I yell and scream at my family.	.49

D. RESULTS

Loadings above .40 will be considered in interpreting the factors (Table 1).

Factor I: Physical Complaints. This factor indicates a pattern of physical-somatic complaints, including discomforts involving arms, hips, ears, eyes, total body, thinking, etc. Such a pattern of physical complaints has also been reported by Hamilton and White (6), Lorr, O'Connor, and Stafford (12), O'Connor and Stefic (16), Wittenborn (21, 22), and Wittenborn and Holzberg (24).

Factor II: Physical Debilitation. This factor indicates a pattern of physical malfunctionings, somatic incapabilities, bodily inadequacies, loss of physical functionings, and bodily weaknesses. Hamilton and White (6) have isolated an analogous factor.

Factor III: Alcoholism. This factor indicates a pattern of complaints about alcoholism, a straightforward concern with excessive drinking.

Factor IV: Interpersonal Difficulties. This factor indicates a pattern of interpersonal difficulties, including interpersonal rejection, vocation inadequacy, social incapability, and interpersonal verbal problems. Such a factor is tangentially related to social withdrawal (1, 4, 12).

Factor V: Marital Discord. This factor indicates a pattern of marital discord and emotional lability, including complaints about the wife, tearfulness, aggression. Elements of this pattern, relating to depression-helplessness-inadequacy, have been reported by other investigators (1, 10, 11, 12, 22, 23, 24).

Factor VI: Aggression. This factor indicates a pattern of anger, irritation, hostility, violent impulses, and overt aggression. Other studies have reported similar patterns (2, 4, 9, 10, 11, 22).

E. DISCUSSION

The results indicate that the self-reported, behavioral problems-symptoms-complaints of hospitalized psychiatric patients occur in identifiable patterns. Two of the patterns especially (Physical Complaints and Aggression) have emerged in other studies of similar populations. However, a comparison of the authors' results with those from a traditional psychiatric approach indicates that the present patterns are not at all readily converted or translated into the conventional psychiatric model. None of the authors' patterns neatly corresponds to schizophrenic excitement (11, 12, 21, 22, 23, 24), paranoid schizophrenia (4, 10, 11, 12, 21, 22, 23, 24), psychosis (2, 19), phobic compulsivity (22, 23, 24), agitated manic state (10, 12, 21, 22, 23, 24), hebephrenic negativism and resistance (10, 12, 22, 23), conceptual disorganization and impairment (2, 9, 10, 11, 12, 22, 23), perceptual distortion (10, 11, 12), grandiose expansiveness (9, 11), personality disorganization (4), or anxiety (21, 22, 23, 24).

The challenging implication is that a goal-directed approach reveals a significantly different set of behavioral symptom patterns, as compared with a traditional psychiatric diagnostic approach. In the traditional psychiatric approach, the symptom patterns signify component elements of psychiatric diagnostic entities and the parameters of "mental illness." Within a goal-directed approach, the significance of the patterns lies in their functional utility as the major behavioral avenues toward gaining hospital admission and toward achieving psychological goals within the hospital setting.

Accordingly, the results suggest that the authors' sample of hospitalized, male, psychiatric patients is characterized by the following major patterns of behavioral problems-symptoms-complaints: physical complaints, physical debilitation, alcoholism, interpersonal difficulties, marital discord, and aggression. As a working hypothesis for further research, it is suggested that these patterns constitute the major behavioral avenues by which male psychiatric patients present themselves for hospital admission, and the major behavioral avenues toward achieving psychological goals available within a hospital setting.

Understood in terms of goal-directed behavioral avenues, the six patterns offer a basis for categorizing patients in terms of manifest, behavioral problems-symptoms-complaints. It is suggested that these six categories constitute a tentative foundation for goal-directed treatment programs targeted upon the six patterns of psychological problems-symptoms-complaints. For example, treatment programs might focus upon problems of marital discord rather than hebephrenic schizophrenia, paranoid personality, or other psychiatric illnesses; interpersonal difficulties might utilize group therapy techniques to a greater extent than would problems centering on physical complaints. The implication is that behavioral categories serve as a working functional taxonomy for purposes of treatment programming and prediction.

F. SUMMARY AND CONCLUSIONS

This study was designed to determine whether the behavioral problems-symptoms-complaints of hospitalized psychiatric patients occur in identifiable patternings which, rather than fitting into the current psychiatric diagnostic system, represent major behavioral avenues toward achieving hospitalization and psychological goals within a psychiatric setting. The present study stems directly from an approach in which symptomatology is described in terms of goal-directed psychological functions.

The *S*s were 302 hospitalized male psychiatric patients. Data consisted of responses to an individually administered, self-report checklist of 150 psychological problems-symptoms-complaints. With the use of a root-one criterion on 84 items of sufficiently high item variance, six factors were extracted.

The results indicate that the self-reported, behavioral problems-symptoms-complaints of hospitalized male psychiatric patients occur in the following patterns: I. Physical Complaints; II. Physical Debilitation; III. Alcoholism; IV. Interpersonal Difficulties; V. Marital Discord; and VI. Aggression. A comparison of these factor-patterns with those from the traditional psychiatric approach indicates that the authors' patterns are not readily converted or translated into the conventional psychiatric model. As a working hypothesis for further research, the results suggest that the six factor-patterns constitute the major behavioral avenues by which male psychiatric patients present themselves for hospital admission, and the major behavioral avenues toward achieving psychological goals within the hospital setting. The implication is that these behavioral categories serve as a working functional taxonomy for purposes of treatment programming and prediction.

REFERENCES

1. ANKER, J. M. A note on the factor structure of the neuropsychiatric chronicity scale. *J. Consult. Psychol.*, 1962, **26**, 198.
2. BRYANT, J. H., WURSTER, C. R., HINE, F. R., & DAWSON, J. G. A factorial analysis of behavior ratings of hospitalized mental patients. *J. of Psychol.*, 1958, **46**, 167-173.
3. DEGAN, J. W. Dimensions of a functional psychosis. *Psychomet. Monog.*, 1952, **6**, 1-41.
4. GUERTIN, W. H. A factor-analytic study of schizophrenic symptoms. *J. Consult. Psychol.*, 1952, **16**, 308-312.
5. HALEY, J. Strategies of Psychotherapy. New York: Grune & Stratton, 1963.
6. HAMILTON, J., & WHITE, J. M. Clinical syndromes in depressive states. *J. Ment. Sci.*, 1959, **105**, 987-998.
7. HORN, J. L. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 1965, **30**, 179-185.
8. KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187-200.
9. LORR, M. The Wittenborn psychiatric syndromes: An oblique rotation. *J. Consult. Psychol.*, 1957, **21**, 439-444.
10. LORR, M., JENKINS, R. L., & O'CONNOR, J. P. Factors descriptive of psychopathology and behavior of hospitalized psychotics. *J. Abn. & Soc. Psychol.*, 1955, **50**, 78-86.
11. LORR, M., KLETT, C. J., & MCNAIR, D. M. Syndromes of Psychosis. New York: Macmillan, 1963.
12. LORR, M., O'CONNOR, J. P., & STAFFORD, J. W. Confirmation of nine psychotic symptom patterns. *J. Clin. Psychol.*, 1957, **13**, 252-257.
13. MAHRER, A. R. The psychodynamics of psychiatric hospitalization. *J. Nerv. & Ment. Dis.*, 1962, **135**, 354-360.

14. ———. The goals of intensive psychotherapy. In A. R. Mahrer (Ed.), *The Goals of Psychotherapy*. New York: Appleton-Century-Crofts, 1967. Pp. 270-287.
15. ———. The psychological problem inventory. *Psychol. Rep.*, 1967, **20**, 711-714.
16. O'CONNOR, J. P., & STEFIC, E. C. Some patterns of hypochondriasis. *Educ. Psychol. Meas.*, 1959, **19**, 363-371.
17. SCHWARTZ, C. G., SCHWARTZ, M. S., & STAUNTON, A. H. A study of need-fulfillment on a mental hospital ward. *Psychiatry*, 1951, **14**, 223-242.
18. SZASZ, T. S. The Myth of Mental Illness. New York: Hoeber-Harper, 1961.
19. TROUTON, D. S., & MAXWELL, A. E. The relation between neurosis and psychosis. *J. Ment. Sci.*, 1956, **102**, 1-21.
20. WANKLIN, J. M., FLEMING, D. F., BUCK, C. W., & HOBBS, G. E. Factors influencing the rate of first admission to mental hospital. *J. Nerv. & Ment. Dis.*, 1966, **21**, 103-116.
21. WITTENBORN, J. R. Symptom patterns in a group of mental hospital patients. *J. Consult. Psychol.*, 1951, **15**, 290-302.
22. ———. The dimensions of psychosis. *J. Nerv. & Ment. Dis.*, 1962, **134**, 117-128.
23. ———. Distinctions within psychotic dimensions: A principal component analysis. *J. Nerv. & Ment. Dis.*, 1963, **137**, 543-547.
24. WITTENBORN, J. R., & HOLZBERG, J. D. The generality of psychiatric syndromes. *J. Consult. Psychol.*, 1951, **15**, 372-380.

Department of Psychology
Miami University
Oxford, Ohio 45056

Published as a separate and in *The Journal of Psychology,* 1968, **68**, 159-163.

FORM EQUIVALENCE OF THE REVISED VISUAL RETENTION TEST IN MENTAL RETARDATES: AN EXTENSION*[1]

Department of Psychology, University of Houston

JAMES A. RICE

A. INTRODUCTION

In a recently published study, Benton *et al.* (2) found Forms C and D of the Benton Visual Retention Test, VRT (1), to yield different results in a sample of 60 children, ages 5-11, *IQ*'s 85-115, on the direct copying administration. Form D proved to be somewhat easier than Form C as a copying task. Collection of data on 81 additional children led to some attenuation of this difference, but it still persisted. However, Brown and Rice (3), using the data for 80 low *IQ* children (mean *IQ* 80), found both Forms C and D to yield results not statistically different on both memory and direct copying administrations.

The VRT, like the great majority of instruments designed to assess some aspect of intellectual functioning, generally finds its greatest utility in populations deficient in various aspects of intellectual ability. The results obtained with an intellectually subnormal population are not necessarily consistent with results obtained on an intellectually normal population. This fact, together with a need for an additional form from time to time either for research purposes or for the reason that an administration is occasionally invalid, constitutes the basis for the present study, an extension of a previous investigation by Brown and Rice (3).

It was hypothesized that, for six groups of mentally retarded children matched on chronological age (CA) and mental age (MA), there would be no significant differences in results obtained among Forms C, D, and E or in order of administration of the VRT for each of 20 scoring categories for both memory and direct copying administrations.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 14, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This investigation was supported in part by Research Grant NB-00616 from the National Institute of Neurological Diseases and Blindness. The author wishes to express appreciation to Dr. A. L. Benton for valuable advice; to the Hospital School, Pine School Section, University of Iowa, for providing the opportunity to conduct this study; and to Sue Penney, Ph.D., University of Iowa, and Byron Egeland, Ph.D., Syracuse University, for assisting with the study.

The specific questions to be answered by the A × B design used in the study were: (*a*) is there an interaction between form of the test used and order of administration—memory first or direct copying first? That is, in groups matched for CA and MA, does one form yield significantly different results from another when administered first as a memory task, then as a copying task? If not, then (*b*) is there a significant difference in the results obtained among Forms C, D, and E, and (*c*) does order of administration —memory first or direct copying first—yield significantly different results?

B. PROCEDURE

1. *Subjects*

The subjects were 72 educable mentally retarded children (mean *IQ* 75, *SD* 9.2) enrolled during the 1965-66 school year in an elementary special education program at the Pine School Section, Hospital School, University of Iowa, Iowa City, Iowa. Table 1 presents statistics describing the sample used in this study. Six groups (cells) are indicated as required by the design to be discussed in the next section. The A dimension was the form used: C, D, E, memory administration in the first analysis, and copying in the second. The B dimension was order of administration, either memory first followed by copying or copying first followed by memory administrations.

TABLE 1
MATCHING STATISTICS

		CA		MA		*F*	
Group	*N*	$\bar{x}$	*SD*	$\bar{x}$	*SD*	CA	MA
			Cells: memory administrations				
1	10	118.5	25.2	90.2	17.8		
2	12	110.1	28.2	82.4	20.5		
3	13	110.7	29.9	83.2	21.3		
4	14	118.7	21.7	84.2	20.9		
5	11	116.5	26.5	87.4	19.3		
6	12	110.5	21.4	86.8	19.9	.294	.219
			Cells: copy administrations				
1	13	117.0	29.1	84.6	21.1		
2	11	112.4	29.3	85.3	18.5		
3	11	108.0	25.3	85.0	21.2		
4	14	113.6	24.0	87.4	18.6		
5	12	114.5	21.7	84.8	22.4		
6	11	118.5	24.0	85.5	19.4	.215	.0317

2. *Administration and Scoring*

Graduate students in school psychology with from one to three years training administered the VRT. A graduate student in school psychology with three years training and one year experience in the Des Moines school system scored the protocols.

3. *Instrument*

The Benton Visual Retention Test (VRT) "is designed to assess visual perception, visual memory and visual constructive abilities" (1, p. 1). For each of two administrations—memory after 10-second exposure and direct copying —10 cards are presented the *S* on which are geometric line drawings. The *S* is to copy or to reproduce these designs from memory. Scoring is by number of designs correctly reproduced and by number of errors made. These are divided into six error categories—omissions, distortions, perseverations, rotations, misplacements, and size errors. Total error score is also divided into left errors and right errors indicating that the error occurred in the subject's left or right visual field. This provides for a total of 10 scoring categories, although at this time number correct and total number of errors are the two of principal concern diagnostically.

C. Results and Discussion

Table 2 reports the results of this study. No obtained *F*-value was significant at the .01 level. More liberally, using the .05 level of significance, only two obtained *F*-values are significant, which is less than would be expected by chance for the 60 *F*-values obtained. The specific questions posed in the Introduction may be answered in this manner: (*a*) The finding of no interaction between form and order of administration suggests that in a sample of educable mentally retarded children the effect of order of administration —memory first or direct copying first—does not change significantly the results obtained with one form over those obtained with another form. The finding of no interaction permits answering the two other questions. (*b*) There is no significant difference among Forms C, D, and E of the VRT, and (*c*) there is no significant difference between orders of administration—memory first or direct copying first. These conclusions apply to all scoring categories.

D. Summary

Earlier research suggests that the three forms of the Benton Visual Retention Test (VRT) are not equivalent in normal children. The VRT finds considerable application in mentally retarded populations. The present study embraced this issue in terms of form equivalence and order of administration of the forms. Results suggest that, when used with a sample of educable mental retardates, the VRT may be assumed to consist of three parallel forms applicable as either memory or direct copying tasks in which order of administration is not significant.

TABLE 2

OBTAINED F-VALUES FOR TEST OF MEAN DIFFERENCES OF EQUIVALENCE OF FORMS ANALYSIS FOR VISUAL RETENTION TEST WITH EDUCABLE MENTALLY RETARDED CHILDREN

Variable	No. correct	No. errors	O	D	P	R	M	S	L	Rt
					Memory administration scoring dimensions[a]					
$F_{AB}(2,66) =$	.60	.08	.69	.49	.61	.03	.53	.31	.18	.09
$F_{A}(2,66) =$	.82	.04	.42	1.23	1.30	.42	1.39	.96	2.72	2.07
$F_{B}(1,66) =$	.95	1.11	.01	2.01	.45	.33	2.16	2.15	.56	1.26
					Direct copying administration scoring dimensions[a]					
$F_{AB}(2,66) =$	1.42	.42	.33	.85	1.71	1.67	3.67*	.37	.24	.87
$F_{A}(2,66) =$	2.90	1.53	.08	1.47	1.87	1.88	.40	.91	.96	3.65*
$F_{B}(1,66) =$	.01	.33	1.80	.50	3.44	.02	.24	.11	.48	.04
$.01F(1,60) = 7.08$			$.01F(2,60) = 4.98$			$.05F(1,60) = 4.00$			$.05F(2,60) = 3.15$	

* Significant beyond the .05 level.

[a] Error categories key: O = Omissions, D = Distortions, P = Perseverations, R = Rotations, M = Misplacements, S = Size, L = Left Errors, and Rt = Right Errors.

References

1. Benton, A. L. The Revised Visual Retention Test: Clinical and Experimental Applications. New York: Psychol. Corp., 1963.
2. Benton, A. L., Spreen, O., Fangman, M. W., & Carr, D. L. Visual Retention Test, Administration C: Norms for Children. *J. Spec. Educ.*, 1967, **1**(2), 151-156.
3. Brown, L. F., & Rice, J. A. Form equivalence analysis of the Benton Visual Retention Test in low *IQ* children. *Percept. & Motor Skills*, 1967, **24**, 737-738.

Department of Psychology
University of Houston
Cullen Boulevard
Houston, Texas 77004

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Act of October 23, 1962: Section 4369, Title 39, United States Code)

Publisher: File two copies of this form with your postmaster.

Form Approved, Budget Bureau No. 46-R029

1. DATE OF FILING: Sept. 30, 1967

2. TITLE OF PUBLICATION: The Journal of Psychology

3. FREQUENCY OF ISSUE: 6 issues: published Jan., March, May, July, Sept., Nov.

4. LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, city, county, state, ZIP code)*: The Journal Press, 35 New Street, Worcester, Massachusetts 01605

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*: The Journal Press, 2 Commercial Street, Provincetown, Mass. 02657

6. NAMES AND ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and address)*: The Journal Press, 35 New Street, Worcester, Massachusetts 01605

EDITOR *(Name and address)*: Editorial Board: 2 Commercial Street, Provincetown, Mass. 02657

MANAGING EDITOR *(Name and address)*: Powell Murchison, 2 Commercial Street, Provincetown, Mass. 02657

7. OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)*

NAME	ADDRESS
The Journal Press	2 Commercial St., Provincetown, Mass. 02657
Dorotea P. Murchison	2 Commercial St., Provincetown, Mass. 02657
Powell Murchison	2 Commercial St., Provincetown, Mass. 02657
Marjorie E. Monteith	19 Hollins Dr., Santa Cruz, Calif. 95060

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
Same as above	

9. Paragraphs 7 and 8 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting, also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner. Names and addresses of individuals who are stockholders of a corporation which itself is a stockholder or holder of bonds, mortgages or other securities of the publishing corporation have been included in paragraphs 7 and 8 when the interests of such individuals are equivalent to 1 percent or more of the total amount of the stock or securities of the publishing corporation.

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	SINGLE ISSUE NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED *(Net Press Run)*	1921	2025
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	none	none
2. MAIL SUBSCRIPTIONS	1543	1585
C. TOTAL PAID CIRCULATION	1543	1585
D. FREE DISTRIBUTION *(including samples)* BY MAIL, CARRIER OR OTHER MEANS	10	10
E. TOTAL DISTRIBUTION *(Sum of C and D)*	1553	1595
F. OFFICE USE, LEFT-OVER, UNACCOUNTED, SPOILED AFTER PRINTING	368	430
G. TOTAL *(Sum of E & F—should equal net press run shown in A)*	1921	2025

I certify that the statements made by me above are correct and complete.

(Signature of editor, publisher, business manager, or owner)

Powell Murchison, Business Manager

Journal of Genetic Psychology

Founded by G. Stanley Hall in 1891; reorganized by Carl Murchison in 1925

Quarterly, published in March, June, September, and December. Devoted to child behavior, animal behavior, and the problems of aging. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1891 at $15.00 per volume, plus transportation.

Journal of General Psychology

Founded in 1927 by E. B. Titchener and Carl Murchison

Quarterly, published in January, April, July, and October. Devoted primarily to experimental, theoretical, clinical, and historical psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1927 at $15.00 per volume, plus transportation.

Journal of Social Psychology

Founded in 1929 by John Dewey and Carl Murchison

Bimonthly, beginning with February. National, group, cultural, racial, and differential psychology. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Bimonthly, three volumes per year; per annum, $30.00; per volume, $15.00; single numbers, $7.50. First number is dated February, 1930. Complete sets from 1930 at $15.00 per volume, plus transportation.

Each author will receive 100 unbound reprints gratis. Additional reprints may be purchased at a nominal cost when the author returns page-proof.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Club Rate for Five International Psychological Journals

	PER ANNUM
THE JOURNAL OF GENETIC PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
GENETIC PSYCHOLOGY MONOGRAPHS (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF GENERAL PSYCHOLOGY (Quarterly, two volumes per year)	$20.00
THE JOURNAL OF SOCIAL PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
THE JOURNAL OF PSYCHOLOGY (Bimonthly, three volumes per year)	$30.00
TOTAL	$120.00
TEACHERS AND STUDENTS EDITION, *Restricted Sale*	$50.00

The *Teachers and Students Edition* is open to any individual psychologist, teacher, student, or physician but requires an agreement that the edition will not be offered for sale till after five years from publication and an agreement that the edition will not be substituted for an institutional subscription.

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

Genetic Psychology Monographs

Founded in 1925 by Carl Murchison

Quarterly: February, May, August, November. Two volumes per year. Child behavior, animal behavior, and the problems of aging. Each number one or more complete researches. Manuscripts may be sent to any member of the Editorial Board, or may be sent directly to The Journal Press. All subscriptions and business communications should be sent directly to The Journal Press. Two volumes annually. Per annum, $20.00; per volume, $15.00; single numbers, $7.50. Complete sets from 1926 at $15.00 per volume, plus transportation.

NOTICE. Effective immediately we offer prior publication for certain selected monographs. Charges will be based on the announced rates for *The Journal of Psychology. For a limited period of time our policy is changed for regular monographs: Authors will be charged for cuts and expensive tables, but no page charge will be imposed.* Two hundred reprints of each monograph will be supplied gratis.

The listing of monographs for 1926 through 1963 will be printed in the first half of each volume of all five journals. Monographs published after 1963 will be listed in the second half of each volume.

Genetic Psychology Monographs *(continued)*

VOLUME 69—January-June, 1964

1. Occupational and industrial therapy in the treatment of psychiatric patients: A controlled study of efficacy in an intensive-treatment institution—M. Hundziak and B. Pasamanick
An analysis of Carl R. Rogers' theory of personality—M. S. Krause
Personality aspects of teaching: A predictive study—H. Zimiles, B. Biber, W. Rabinowitz, and L. Hay
Situational and personal antecedents of incipient alienation: An exploratory study—W. H. Jarrett and A. O. Haller
2. Age changes in children's Mosaic responses from five to ten years—L. B. Ames and F. L. Ilg
A study of the interview process: The caseworker-client relationship—H. J. Meyer, E. F. Borgatta, and D. Fanshel
A study of introversion-extroversion as a personality dimension—R. M. Allen, H. M. Richer, and R. J. Plotnik
Psychological testing with Northwest Coast Alaskan Eskimos—C. E. Preston
A study in family pathology—M. B. Hans

VOLUME 70—July-December, 1964

1. Structure and function—H. Gutman
The Lowenfeld Mosaic Test: Norms for five-to-ten-year-old American public-school children and comparative study of three groups—L. B. Ames, F. L. Ilg, and J. August
Social and personality correlates of children's estimates of height—J. P. Shaffer
The development of ethnic attitudes in New Zealand school children—G. M. Vaughan
2. The antecedents of two varieties of vocational interests—L. E. Tyler
Personality in old age—P. E. Slater and H. A. Scarr
Variables affecting experimenter bias in a group situation—R. Rosenthal, C. J. Friedman, C. A. Johnson, K. Fode, T. Schill, R. C. White, and L. L. Vikan
Ordinal position and sex-role identification—B. G. Rosenberg and B. Sutton-Smith
Personality correlates of accuracy of self-appraisal in four-year-old children—Paul S. Weiner

VOLUME 71—January-June, 1965

1. Author index for Genetic Psychology Monographs 1926-1964—L. W. Ferguson
A theoretical, developmental, and clinical investigation of embarrassment—J. M. Sattler
School readiness as evaluated by Gesell developmental, visual, and projective tests—F. L. Ilg, L. B. Ames, and R. J. Apell
Rorschach scales of regression in psychosis—S. Neiger, A. G. Slemon, and D. A. Quirk
The effects of extra tactile stimulation on a group of institutionalized infants—L. Casler
2. Human-eyelid conditioning: The recent experimental literature—L. E. Ross and T. F. Hartman
Social status and schoolroom aggression in third-grade children—R. Toigo
Overt and covert homosexual problems among male college students—L. J. Braaten and C. D. Darling
Role conceptions of the psychiatric aid—M. P. Lawton and A. E. Goldman
Changes in Study of Values scores during the four years of college—W. Huntley

VOLUME 72—July-December, 1965

1. A study of process and method in three-year-old twins and triplets learning to read—W. Fowler
Individualized teaching with therapeutic aims: A methodological study—M. Kohn and M. Rudnick
Ability and social adjustment at midlife of persons earlier judged mentally deficient—E. L. Miller
2. Development of incentive values in childhood—S. L. Witryol, D. J. Tyrrell, and L. M. Lowden
Longitudinal changes in intolerance and authoritarianism for subjects differing in amount of college education over four years—W. T. Plant
An NDEA counseling and guidance institute: Prediction, performance, and follow-up—R. H. Dana
The influence of language on classification: A theoretical model applied to normal, retarded, and deaf children—H. G. Furth and N. A. Milgram
Time estimates as affected by incentive class and motivational level—M. Felix
Clinical psychology: An historical survey—V. S. Sexton

VOLUME 73—January-June, 1966

1. Vulnerability in infants and young children: A pilot study—G. Heider
2. Racial preference and ego-identity among white and Bantu children in the Republic of South Africa—A. J. Gregor and D. A. McPherson
Mother-child interaction in cerebral palsy: Environmental and psychosocial obstacles to cognitive development—E. Shere and R. Kastenbaum
Children's stories—L. B. Ames

VOLUME 74—July-December, 1966

1. Changes in personality for groups completing different amounts of college over two years—W. T. Plant and C. W. Telford
Brain dynamism in perception of reversible figures—G. K. Yacorzynski
Changes in Rorschach response throughout the human life span—L. B. Ames
Mnemonic virtuosity: A study of chess players—A. Binet (Translated by M. L. Simmel and S. B. Barron)
2. The development of behavior patterns in children—B. B. Tyler, F. B. Tyler, and J. E. Rafferty
Correlates of classroom functioning—L. M. Smith and B. B. Hudgins
Children's behavior syndromes and parental responses—R. L. Jenkins, E. NurEddin, and I. Shapiro

VOLUME 75—January-June, 1967

1. Studies of an elementary game of strategy—B. Sutton-Smith and J. M. Roberts
Child care practices anteceding three patterns of preschool behavior—D. Baumrind
The impact of cognitive maturity on the development of sex-role attitudes in the years 4 to 8—L. Kohlberg and E. Zigler
2. Differential factor structures in mathematical ability—P. S. Very
Perception of the vertical as a function of affective stimuli and extraneous stimulation in schizophrenic and normal subjects—J. O. Wells and W. E. Caldwell
Mid-life attainment of the mentally retarded: A longitudinal study—W. R. Baller, D. C. Charles, and E. L. Miller

VOLUME 76—July-December, 1967

1. Scalogram analysis of some developmental sequences of infant behavior as measured by the Bayley Infant Scale of Mental Development—R. Kohen-Raz
The leader and the lost: A case study of indigenous leadership in a poverty program Community Action Committee—L. A. Zurcher
A developmental study of the perception of people—R. N. Wolman
Human fears: A factor analytic study of three age levels—G. W. Russell

VOLUME 77—January-June, 1968

1. The fulfillment of promise: 40-year follow-up of the Terman Gifted Group—M. H. ODEN
   A scale for parents' ratings: Some ipsative and normative correlations—R. N. WALKER
   A methodological investigation of Piaget's concept of conservation of substance—A. C. BITTNER AND MARTIN M. SHINEDLING

THE JOURNAL PRESS
2 Commercial Street
Provincetown, Massachusetts
U. S. A.

PUBLICATION IN THE JOURNAL OF PSYCHOLOGY

The Journal of Psychology provides a unique service to authors. Articles accepted for publication in this periodical appear in print with the least possible amount of delay. In line with this policy, all steps of the acceptance procedure are maximally streamlined, and the author's work is presented with a bare minimum of editing or restriction of any kind. *The Journal of Psychology* is the ideal medium for presentation of timely professional material.

Authors are requested to exercise unusual care in the preparation of papers for this journal. Accuracy, grammar, and style should be checked and rechecked before a clean version of the manuscript is submitted. Careful attention should be paid to the General Instructions on the back cover of each issue and to the announcement on the inside front cover.

Authors are particularly requested to note the fact that all parts of the manuscript—including references, footnotes, and tables—must be double-spaced.

$30.00 per annum
$15.00 per volume
$7.50 per half volume

BIMONTHLY
Three volumes per year
Immediate publication

March, 1968
Volume 68, Second Half

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold

under the restrictions of a bilateral contract, and may not be resold for a period of five years from the date of publication.

MARCH, 1968

(Manuscripts are printed in the order of final acceptance)

Age differences in the relationship between EEG arousal and reaction time . 167
By Larry W. Thompson and Jack Botwinick
Rater reliability of a depressive rating scale 173
By Joseph Gilbert and James E. Gilbert
The effects of anxiety and item difficulty sequence on achievement testing scores 181
By Albert D. Smouse and David C. Munz
Short term memory and coding behavior as a function of set during sequential trials 185
By E. P. Horne and Bruce C. Flanagan
Results of the National Teacher Examinations at Tulane University, 1948-1966 . 193
By E. C. Hunter
Color response to the TAT: An instance of eidetic imagery? 203
By Peter W. Sheehan

(OVER)

Copyright © 1968 by The Journal Press
Provincetown, Massachusetts, U. S. A.
All rights reserved

Published at 35 New St., Worcester, Mass.
Second-class postage paid at Worcester, Mass.
Send all mail to Managing Editor, 2 Commercial Street, Provincetown, Massachusetts

Individual differences in figural aftereffects: Some problems and potentials . 211
By Michael Wertheimer and Fred H. Herring

The language of children in the "inner city" 215
By Theron Alexander, Judith Stoyle, and Charles Kirk

Two scales to measure cognitive rigidity 223
By James G. Snider and Grant C. Drakeford

An analysis of multitrial free recall learning with input-order held constant . 227
By William A. Roberts

Attempted suicide as a hostile act 243
By David Lester

The role of the father in guidance clinic consultations 249
By P. O. Davidson and A. R. Schrag

Programmed instruction, teacher ability, and subject matter difficulty . . . 257
By John J. La Gaipa

A note on the use of judges 261
By Lewis W. Brandt

A psychological analysis of "resistance" fighters during World War II . . 267
By Andre L. Juliard and Alex S. Juliard

Interpersonal attraction as a function of freedom of choice 281
By S. Thomas Friedman and David Sherrill

Use of the Edwards Personal Preference Schedule with alcoholics in a vocational rehabilitation program 287
By Allen Goss, Tom Morosko, and Robert Sheldon

The correlation of two self-disclosure inventories with actual self-disclosure: A validity study 291
By Darhl M. Pedersen and Vincent J. Breglio

Institutionalized retardates' reactions to authority figures: God and monster drawings 299
By Gabriel Pustel, Manny Sternlicht, and Louis Siegel

The MMPI measurement of phenotypic anxiety: I. Exploratory study . . . 305
By Raymond D. Fowler, Jr., F. A. Coyle, Jr., P. C. Reed, and Chester A. White

Transfer of social influence effects on psychophysical judgments 313
By Robert K. Yin and Herbert D. Saltzstein

Judgments of ethnic groups 321
By Joseph C. Finney

Published as a separate and in *The Journal of Psychology,* 1968, 68, 167-172.

AGE DIFFERENCES IN THE RELATIONSHIP BETWEEN EEG AROUSAL AND REACTION TIME*

Duke University Medical Center

Larry W. Thompson[1] and Jack Botwinick[2]

A. Introduction

While the terms "set" and "expectancy" are often used to explain the role of the preparatory interval (PI) in a reaction time (RT) experiment, it has been suggested that their meaning may not be very different from the terms "arousal" and "activation" used in conjunction with indices involving the electroencephalogram, EEG (2). In fact, a series of studies has been carried out for the purpose of trying to relate RT with EEG measures. Conclusions based on these studies remain equivocal; while several pointed to an EEG-RT relationship (6, 7, 9, 11, 12), a failure to find such a relationship has also been frequently reported (1, 7, 8, 13). In similar fashion, the concepts of "set" or "expectancy" have been introduced to account for any age differences in RT which occur as a result of variations in PI (3, 4). Efforts to relate RT and EEG measures in the elderly have yielded equally equivocal results. Birren (1), for example, found no relationship between EEG frequency or voltage measures and RT, but Surwillo (12) demonstrated that the mean period of the EEG in the elderly could account for much of the variance of the RT. Neither of these two studies employed warning signals to manipulate set prior to the onset of the stimulus to respond.

The present study was designed to explore the manner in which an EEG activation pattern may relate both to RT and to the experimental condition of "set" or "expectancy," as reflected in manipulations of PI durations and series. In addition, these functions were investigated with respect to age of *S*.

B. Method

1. *Subjects*

The *S*s of this study constituted a subsample of *S*s described elsewhere (5).

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 20, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was supported in part by P.H.S. Research Grant HD-00668 from the National Institutes of Health.

[2] This study was also supported in part by a P.H.S. Research Career Program Award (5153) from the Child Health and Human Development Institute and by P.H.S. Research Grant HD-01325 from the National Institutes of Health.

This subsample of white community residents included two age groups: elderly (62-87 years) and young adults (19-35 years). The elderly *S*s had 12 or more years of formal education and were relatively healthy, with mild complaints of the usual chronic illnesses. In addition to health and education, the criteria for selection included the presence of measurable alpha activity in the parieto-occipital tracing of the EEG in order that a quantitative analysis of this measure could be obtained using manual techniques. There were 16 men, with ages ranging from 68-87 years ($\bar{x} = 76.7$), and 10 women from 67-87 years ($\bar{x} = 78.5$). The same number of men and women *S*s constituted the young group. The men were aged 19-32 years ($\bar{x} = 23.0$) and the women were aged 19-35 ($\bar{x} = 23.8$). The young group was comprised of students, secretarial staff, and housewives.

2. *Procedure*

The procedure has been described in detail in a previous study (13), and only the important features are described here. Briefly, *S* initiated each RT trial by pressing a telegraph key. Approximately two seconds after the key was pressed, a warning tone of .5-second duration (400 cps, 75 db intensity) was presented. The PI was indexed by an auditory stimulus (1000 cps, 85 db) which was terminated by the finger lift response. PIs were .5-, 3.0-, 6.0-, and 15.0-second duration in both a regular and irregular series. In the regular series each PI was constant for a given block of trials, whereas in the irregular series PIs were varied in a prearranged manner, in such a way that each PI duration preceded all other PIs an equal number of times.

EEG tracings were recorded simultaneously from the left parieto-occipital area during the middle RT trials of each of the different PI conditions. In all, EEGs were recorded during 82 RT trials for each *S*. A resting control record was also obtained. The tracings were recorded by a Grass Model III EEG which was set at increased paper speed (6 cm per second) and increased gains (1 cm = 50 microvolts). Electrode placements were 1.5-cm left of the midline and 5- and 12-cm, respectively, above the occiput.

The EEG measure in the present experiment consisted of the relative change in EEG amplitude from a rest control sample to a .5-second sample associated with each stimulus presentation. The change in EEG amplitude was expressed as a percentage of level during rest control (rest control-stimulus—rest control): thus, the larger the number, the greater the decrease in amplitude relative to the rest control levels. Since alpha blocking (desynchronization) is related to EEG amplitude, it was assumed that this change measure provided a rough index of EEG activation at any given point.

This EEG index of activation was computed by a peak-to-peak amplitude analysis of three 5-second samples of rest control, selected at 30-second intervals and the .5-second sample associated with each stimulus presentation. This involved the .4-second preceding the onset of the stimulus and .1-second overlapping it. The rising surface of every wave in the selected samples, which had an amplitude greater than 10 microvolts, was measured. This measurement was then accumulated for each wave in the selected portion, which gave an estimate of the cumulative peak-to-peak amplitude. It has been shown that desynchronization of the EEG is reflected in a decrease of this measure (13).

C. Results

The EEG measure of activation in relation to the PI is presented in Figure 1 for both old and young *S*s. The data underlying this figure were subjected to a Lindquist Type III analysis of variance (10) and the results may be seen in Table 1. This table shows that, while the two age groups did not differ with respect to overall activation pattern (EEG amplitude change), they did differ in this pattern in relation to the PI; the age-PI interaction was statistically significant at the .01 level in the irregular series, and at the .05 level in the regular series. Figure 1 shows that in the regular series the greatest change in EEG amplitude occurred during the .5-second PI for both age groups of *S*s ($p < .05$). The EEG change for the old group was greater than for the young, but the difference between the two age groups diminished as the PI durations increased. In the irregular series, on the other hand, Figure 1 shows an actual cross-over effect with age; for old *S*s, the maximum EEG change occurred with the shortest PI (.5-second), while for young *S*s the minimum change occurred with this PI.

With respect to RT, Table 1 shows that the old *S*s were significantly slower than the young *S*s ($p < .001$), but the PI did not have a differential effect on the two age groups. That is, unlike previous studies (3, 4) which demonstrate an age-PI interaction with RT measures, no such interaction was found with these data. Rather, the RT relationship to PI for the two age groups emerged as two nearly parallel functions. These RT data are considered in part in a previous study (5). The lack of an expected age by PI interaction effect relative to RT may be attributable to the selection criteria of health, education, and percentage of alpha activity as well as the limited number of RT measures available in each PI.

Independent of age, the PI had an effect on both the EEG and RT measures in the regular series ($p < .05$ and .001, respectively), but not in the irregular series ($p > .05$). Such findings raise the question of a relationship between

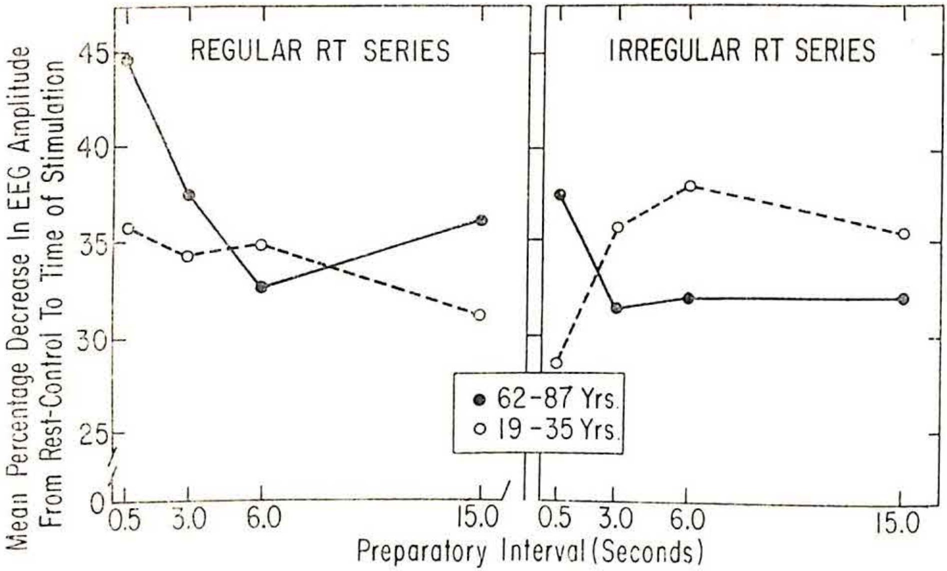


FIGURE 1

Mean Percentage Change in EEG Amplitude as a Function of the PI Duration in Both Regular and Irregular Series

The change in EEG amplitude is from periods of rest-control to .5-second periods involving the reaction time (RT) stimulus (.4-second preceding and .1-second overlapping it). The PI is that duration of time between a warning signal and the RT stimulus.

TABLE 1

Analyses of Variance of EEG Change and RT for Different Preparatory Intervals (PI) in Regular and Irregular Series

		F ratios			
		EEG change		RT	
Source	*df*	Reg.	Irreg.	Reg.	Irreg.
Between *S*s	51				
Age	1	.50	.01	17.30***	7.67***
Sex	1	2.41	3.00	3.89	1.38
Age × sex	1	.93	.13	4.37	4.27*
Error (MS)	48	(.14)	(.14)	(11.96)	(12.99)
Within *S*s	156				
PI	3	3.48*	1.91	25.18***	2.54
Age × PI	3	2.68*	4.04**	1.13	.08
Sex × PI	3	2.57	.50	1.75	.04
Age × sex × PI	3	1.55	.65	.28	1.07
Error (MS)	144	(.61)	(.01)	(2.41)	(1.77)

Note: EEG change refers to the relative difference in summated amplitude between stimulus and rest control levels (rest control-stimulus/rest control).

* $p < .05$.

** $p < .01$.

*** $p < .001$.

EEG activation and RT, but the results do not support this. Intra-*S* correlations were computed between RT and EEG change for each PI in both the regular and irregular series and none of the coefficients even approached statistical significance ($p > .05$); none was large by any criterion; the largest was .21 and most were less than .10. Between-*S* correlations were also obtained for each sex and age group separately, and again no relationship between the two measures was apparent.

D. DISCUSSION

Evidence was presented to show that old and young *S*s did not differ in overall level of EEG change, but did react differently to variations in the PI. If one accepts this EEG measure as a rough index of CNS arousal, the inference can be made that patterns of arousal are different for young and old, depending upon differing states of "set" or "expectancy," as reflected by the experimental manipulation of PI duration. These age differences in arousal patterns are complex and are not easily described by any simple, systematic function.

This age difference in arousal pattern cannot be taken as an explanation of the age differences in RT. The present results show neither an RT-EEG relationship nor overall age differences in EEG change. Thus, the data have left unanswered the question of the variance associated with the slowing of response speed in the aged.

E. SUMMARY

An EEG measure of arousal was investigated in relation to the preparatory interval (PI) in the context of a reaction time (RT) study for each of two age groups. The two groups (19-35 and 62-87 years) reacted differently to the PI variations with respect to the arousal measure, but not with respect to RT. Measures of arousal and RT were uncorrelated. These results, plus the finding of a statistically significant age difference in RT but not in the EEG measure, lead to the conclusion that, in this study at least, EEG changes did not explain the slowing in old age.

REFERENCES

1. BIRREN, J. E. Age changes in speed of behavior: Its central nature and physiological correlates. In A. T. Welford & J. E. Birren (Eds.), *Behavior, Aging, and the Nervous System*. Springfield, Ill.: Thomas, 1965. Pp. 191-216.
2. BOTWINICK, J. Theories of antecedent conditions of speed of response. In A. T. Welford & J. E. Birren (Eds.), *Behavior, Aging, and the Nervous System*. Springfield, Ill.: Thomas, 1965. Pp. 67-87.
3. BOTWINICK, J., & BRINLEY, J. F. Aspects of RT set during brief intervals in relationship to age and sex. *J. Geront.*, 1962, **17**, 295-301.

4. Botwinick, J., & Thompson, L. W. Components of reaction time in relation to age and sex. *J. Genet. Psychol.*, 1966, **108**, 175-183.
5. Botwinick, J., Brinley, J. F., & Birren, J. E. Set in relation to age. *J. Geront.*, 1957, **12**, 300-305.
6. Dustman, R. E., Boswell, R. S., & Porter, P. B. Beta brain waves as an index of alertness. *Science*, 1962, **137**, 533-534.
7. Fedio, P., Mirsky, A. F., Smith, W. J., & Parry, D. Reaction time and EEG activation in normal and schizophrenic subjects. *Electroenceph. Clin. Neurophysiol.*, 1961, **13**, 923-926.
8. Hermelin, B. M., & Venables, P. H. Reaction time and alpha blocking in normal and severely subnormal subjects. *J. Exper. Psychol.*, 1964, **67**, 365-372.
9. Lansing, R. W., Schwartz, E., & Lindsley, D. B. Reaction time and EEG activation under alerted and nonalerted conditions. *J. Exper. Psychol.*, 1959, **58**, 1-7.
10. Lindquist, E. F. Design and Analysis of Experiments in Psychology and Education. Boston, Mass.: Houghton Mifflin, 1953.
11. Monnier, M. Retinal, cortical and motor responses to photic stimulation in man: Retino-cortical time and opto-motor integration time. *J. Neurophysiol.*, 1952, **15**, 469-486.
12. Surwillo, W. W. The relation of simple response time to brain-wave frequency and the effects of age. *Electroenceph. Clin. Neurophysiol.*, 1963, **15**, 105-114.
13. Thompson, L. W., & Botwinick, J. The role of the preparatory interval in the relationship between EEG alpha blocking and reaction time. *Psychophysiol.*, 1966, **3**, 131-142.

Department of Psychiatry
Duke University Medical Center
Durham, North Carolina 27706

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 173-180.

RATER RELIABILITY OF A DEPRESSIVE RATING SCALE*

VA Hospital, Charleston, South Carolina; and VA Hospital, Gulfport, Mississippi

JOSEPH GILBERT AND JAMES E. GILBERT

A. INTRODUCTION

The investigation reported by this article is felt to be conceivably of some significance in four distinct ways: (*a*) in terms of the instrument employed (the Gilbert Depressive Rating Scale, which is constructed to yield a quantitative measure of depression in any nonpsychotic patient population); (*b*) with respect to the use of *independent* ratings simultaneously obtained in a group setting; (*c*) relative to a statistical evaluation assessing the reliability of ratings obtained by a straightforward analysis of variance procedure rather than classical correlation techniques; and (*d*) with relation to the method of assessing the reliability of a measuring instrument by determining the reliability of the persons making the measurement—the extent to which the raters were in accord. The considerations involved in the selection of the experimental design may have some relevance to the problems encountered by others in the evaluation of similar devices in the clinical setting.

The Depressive Rating Scale (DRS) was devised by the second author. It consists of 47 items, all suggestive of a depressive condition of a kind which might be responsive to medication or to electroshock therapy. The scale is divided into three sections; the first consists of symptoms reported by the patient, the second of symptoms observed by the doctor, and the third of symptoms reported by the spouse or appropriate relative of the patient. The symptoms to be rated are scaled from one to five, ranging from "not perceptible or observed" to "markedly present as reported or observed." Examples of the items reported *by the patient* (20 items) are "Feelings of Hopelessness," "Feels Angry," "Impairment of Memory," and "Somatic Complaints." Examples of items observed *by the raters* (nine items) are "Anxiety" (as indicated by such symptoms as sweating, pallor, and tremor of extremities), "Appears Suspicious," "Contact with Reality," and "Expresses Feelings of Jealousy." Examples of the items to be reported *by the spouse* or appropriate relative (18 items) are "Feels Depressed," "Attempts at Suicide," "Difficulty in

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 22, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

Thinking," and "Date of First Attack." The cumulative total of the numbers comprises the Depressive score.

The first factor to be considered in the evaluation of this scale was its reliability. To determine the reliability of the scale by determining the reliability of the raters themselves was the first task involved in the evaluation of the DRS.

Each rater made his rating without knowledge of how the other raters performed. The decision in this instance was that the *independent* ratings would nonetheless be obtained in *the group setting,* with all raters present and rating simultaneously, to avoid repetition and the stereotypy of response which might reasonably be expected to accompany it. A single interview could be anticipated to produce rapport and spontaneity. Hence the patients were seen by all five raters at the same time, but the raters rated the items on the scale entirely independently and without any consultation or discussion among themselves before the ratings were made. In this manner, independent ratings were obtained in the group setting.

The results were analyzed by an analysis of variance technique, the Ebel method (1, pp. 395-397), using the formula

$$\bar{r}ll = \frac{Vp - Ve}{Vp + (K - 1)\,Ve}$$

which is the mean reliability for one rater, and the formula

$$rkk = \frac{Vp - Ve}{Vp}$$

for the reliability for mean ratings from all raters involved. In the above formulas $\bar{r}ll$ is the reliability of ratings for a single rater, Vp is the variance for persons, Ve is the variance for error, and k is the number of raters.

The analysis of variance technique was used, despite the much greater arithmetical computation needed, because virtually every other reliability measure is not appropriate for rating scale data for one reason or another, and because violation of the theoretical assumptions underlying other statistical measures of reliability results in an overestimation of reliability, while if the assumptions of analysis of variance are violated reliability is underestimated, thus loading the probabilities against a higher reliability to that extent. Also it has been well established that analysis of variance and comparable techniques are much less sensitive to violations of assumptions than are other statistical devices.

In the concluding paragraphs of this paper, the results of the analysis of the reliability of the Depressive Rating Scale will be reported. In the material immediately following, the procedures employed for the evaluation of the reliability of the scale will be presented, together with their rationale, for the interest which the problems anticipated and the means by which they were surmounted may have experimentally.

B. PROCEDURE

From two to three patients were seen on one night each week after the closing hours of a Mental Hygiene Clinic in which the patients were to be seen for treatment.[1] The time (5:15 on a Thursday afternoon) was deliberately selected for several reasons: (*a*) it was a convenient time, presumably being after hours of employment for most of our patients; (*b*) it did not interfere with the working hours of the clinic; (*c*) the probable fatigue of the raters, all of whom were clinic personnel, might be expected to operate against the reliability of the scale, making positive results perhaps even more significant than they would be otherwise; (*d*) on the other hand, the patients might be expected to be somewhat fatigued, so that the depressive symptoms should be more clear-cut than they might otherwise be. However, if the items of the Depressive Rating Scale were not reliable this factor should not make a great deal of difference, perhaps less difference than item *c*.

The raters represented a variety of professional disciplines, and consisted of two psychiatric residents, two psychiatric social workers (each with a divergent theoretical orientation: one psychoanalytic, one functional), and a clinical psychologist (the senior author, who was coordinator of the project).[2]

The patients rated were persons seen in the clinic, who were considered by the director of the clinic to be depressed. Not all patients were accompanied by their spouses, but whenever the spouse was available each patient and each spouse were seen individually in the presence of the raters. The latter took turns at asking the patient about the items on the Depressive Rating Scale, although the majority of the questionning (about 70 per cent) was conducted by the male psychiatric resident. The remaining psychiatric resident and two social workers were women. (It was regarded as desirable that both men and women take part in the rating, so that any sex differences involved in rater performances could be observed. As will be indicated, variability between raters was not great, but where variability did occur it was apparent in the statistical data that the women differed more among themselves than they differed from the men.)

---

[1] Richland County Mental Hygiene Clinic, of which the senior author was at the time chief psychologist, and Dr. James E. Gilbert was director.

[2] Besides the senior author, the raters participating in the project were Miss Dimitra George and Miss Atha Cooper (social workers), and Dr. W. S. McClay and Dr. Betty Gerry (psychiatric residents).

It was intended that each patient would receive the Minnesota Multiphasic Inventory (MMPI), and that a separate diagnosis would be made of diagnoses other than depressed. However, since the real object of this initial evaluation was reliability, it was decided that the above objectives were more appropriate to a later evaluation of validity, or the degree to which the scale actually successfully predicted a depressive diagnosis. This decision was influenced as well by the realistic circumstance, very soon encountered, that by the time the patients were seen they had been on antidepressive medication (Nardil and Librium) for long enough to be no longer depressed. Since the purpose of the evaluation was, again, to determine reliability or agreement among the raters on the Depressive Rating Scale, the diagnosis made, the authors recognized, actually not the slightest difference except thereby presumably increasing variability and accordingly operating against consistency between raters. With this realization, and with subsequent difficulty in finding a sufficient number of depressed patients, the investigators added new patients to the rating project at random. Approximately 90 per cent of these cases were known to only one member of the clinic personnel in intake procedure before rating occurred, so that the patients rated were strangers to four or more raters on any one occasion.

Under the circumstances the investigators had anticipated reliabilities in the range of perhaps the 50's or 60's for the DRS if they were fortunate, a quite satisfactory level for a rating scale. Instead, for five raters and 15 subjects the reliability of single raters was .905 and the mean reliability of all five raters was .979 out of a possible 1.00 for total scores. For Category A of the Depressive Rating Scale (ratings of the subjects' replies to the examiner), the reliability of five raters and 15 subjects was .43 for one rater and the mean reliability for all raters was .79. (Of course reliability increases up to a point with increasing length and decreases with decreasing length.) For Category B of the Depressive Rating Scale (observation by the rater of the subject), five raters and 14 subjects, despite the great brevity of this category (about half of Category A) the reliability of a single rater was .55 and the mean reliability of all raters was .82. For Category C (Rating of the spouse's statement about the patient, a section about the length of Category A) five raters and 11 subjects yielded a reliability (despite the reduction in number of subjects) of .85 for one rater and a mean reliability for all raters of .97.

The raters were instructed to rate the severity of what subjects said were their present symptoms on Category A and what spouses said were the severity of subjects' symptoms on Category B. The raters' observations and spouses' statements were much more reliably judged than were the subjects' statements about themselves. It is also of interest that a number of the subjects were considered hysterical individuals who exaggerated their symptoms. The differences between categories raise the possibility of developing a discrepancy score from the categories of the scale which will make it possible to determine pragmatically at least approximately the authenticity of the patient's symptoms.

C. Results

Table 1 summarizes the results of the reliability determination.

The Depressive Rating Scale is thus demonstrated to be of high reliability, especially the total score, and the possibility is raised of a "discrepancy score" being established which might further enhance the significance of the scale.

TABLE 1
Reliability of the Depressive Rating Scale

Category	No. subjects	Mean—All raters	Mean—One rater
Total	15	.979	.905
A	15	.79	.43
B	14	.82	.55
C	11	.97	.85

This leaves the validity of the Depressive Rating Scale essentially undetermined (and it remains a fact that a measuring device can be highly reliable in giving the same results each time without being in the least valid). The reliability of the device, however, seems excellent, at least insofar as the agreement of raters is concerned. (The patient's state would have to remain essentially static if he is to be rated twice, which would not actually be in accord with the purpose of the scale in measuring level of depression at a particular time as opposed to the duration of the depression, which would be involved in obtaining re-ratings of a patient on the scale.) The reliability levels obtained are especially surprising in view of the relatively brief sections of the scale, the fact that an analysis of variance technique was used which underestimates reliability if its assumptions are violated, the rather narrow range of components known to be involved in depression which the scale samples, and the lack of standardized statements on the scale for presentation of its items.

Purely as a means of determining, prior to a properly controlled validity study, whether or not a change in the direction of improved scores on certain psychometric tests tends to be accompanied by improved (reduced scores) on

TABLE 2

CHANGE IN TEST SCORE AND RATING FOLLOWING NARDIL AND LIBRIUM MEDICATION

Category	Digit Span		Similarities		Pic. Compl.		DRS		Color-Nam.	
	Total	Mean	Total	Mean	Total	Mean	Total	Mean	Total	Mean
Pre-medication	70	11.6	81	13.5	63	10.5	566	94.3		31.17 sec.
Post-medication	71	11.83	85	14.7	68	11.33	537	89.5		28.0 sec.

the Depressive Rating Scale, six patients were administered the Digit Span, Similarities, and Picture Completion subtests of the Wechsler, together with the Color-Naming test, before and after receiving Nardil and Librium medication. It is recognized that the number of cases is too small to make any valid deductions possible, even leaving out of consideration the lack of a control group and other essential experimental controls, including those for expectancy, that would be provided in a properly designed validity study. But in spite of this the results presented in Table 2 are suggestive enough to warrant mention here for their very consistency, if for no other reason.

In other words, as scores on the Wechsler subtest go up, scores on the Depressive Rating Scale (DRS) and Color-Naming go down, just as they would be expected to do if the medication is effective. With only six cases, so small a sample as to produce too much variability for real consistency to be expected, in spite of all the uncontrolled variables operating for positive change (and some against it), every test used across the board shows a change in the predicted direction, producing complete uniformity of trend for the group of six as a whole, even though the differences as might well be anticipated are very small. Certainly it does suggest that properly controlled studies of the validity of the Depressive Rating Scale, or a comparable device free from some of the apparent current defects of the DRS, might conceivably yield a practicable technique for measurement of depressive tendencies which could be of considerable potential value in the clinical situation.

D. SUMMARY

The Depressive Rating Scale of James E. Gilbert is described. The reliability of total scores on the Depressive Rating Scale for five raters and 15 subjects was determined to be .905 for one rater and .979 out of a possible 1.00 for the mean or reliability of all raters. Certain procedures were employed to deal with the problems encountered in the assessment of this type of psychometric device (use of independent ratings simultaneously obtained in a group setting, determination of reliability by an analysis of variance technique, and the assessment of reliability by evaluation of the consistency of raters) which may conceivably be applicable to the evaluation by others of similar devices. No systematic, controlled effort was made to evaluate validity at the time of this investigation, but a preliminary appraisal of the absolute change in Wechsler subtest and Color-Naming scores was uniformly associated with DRS scores in the hypothesized direction, although the minuscule number of cases makes this result no more than suggestive that further exploration of the

validity of this, or a similar device, for assessing depressive trends, could be potentially worthwhile.

REFERENCE

1. GUILFORD, J. P. Psychometric Methods (2nd ed.). New York: McGraw-Hill, 1954.

1313 S. Edgewater Drive
Charleston, South Carolina 29407

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 181-184.

THE EFFECTS OF ANXIETY AND ITEM DIFFICULTY SEQUENCE ON ACHIEVEMENT TESTING SCORES*

College of Education and Department of Psychology, The University of Oklahoma

ALBERT D. SMOUSE AND DAVID C. MUNZ

A. INTRODUCTION

Authors writing in the field of test construction have consistently recommended the practice of arranging test items in order of increasing difficulty: i.e., the easier items first followed by the progressively more difficult items (4, 7, 8, 11, 12). Numerous logical arguments have been advocated for and against such a practice. Until recently there has been little evidence for any one argument (9, 13) and considerably less evidence for the practice itself (2).

One major argument supporting this practice has been that students who received easier items early in the test would adapt more rapidly to their test-taking anxiety, thereby increasing their performance, than if the items were ordered with the difficult items at the first of the test and easier items at the end. Some authors have also suggested that a hard-to-easy sequence would produce lower performance scores due to the disruptive effect of encountering difficult items early in the test. Whether any difference in performance on the easy-to-hard sequence (E-H) as compared to the hard-to-easy sequence (H-E) or a random (R) sequence is attributable to the facilitation of test-taking anxiety adaptation on the E-H test, or to a disruptive effect of the H-E sequence, or to both is unknown. The purpose of this study was to evaluate the effect of three item difficulty orders (E-H, H-E, and R) on achievement performance scores in the classroom. The three item difficulty orders were designed and administered in such a way as to detect and assess possible disruptive or adaptive effects on achievement test performance.

B. METHOD

The *S*s were 113 male and female students comprising two sections of the introductory psychology course at the University of Oklahoma. Both sections had the same instructor; however, neither section was aware of the research being conducted. For the final examination of the course, *S*s from the combined sections were randomly divided into two groups, a High test-taking Anxiety

---

* Received in the Editorial Office, Provincetown, Massachusetts, on November 24, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

Group (HAG) which received anxiety-provoking information just prior to the test, and a Normal Anxiety Group (NAG) in which a normal test-taking atmosphere was maintained. Within each group, *S*s were randomly assigned one of three examination forms, described below. Both groups had as much time as needed to finish the test.

The anxiety-provoking treatment consisted of informing the *S*s that because of "widespread cheating" on previous examinations, should their individual performances drop significantly below previous examination scores, they would have to take a special oral examination administered by the coordinator of the introductory psychology sections. Further, the examination was administered by a professor rather than the graduate instructor and was proctored by assistants who continually circulated among the *S*s. [For a comprehensive explanation of the anxiety-provoking procedure and discussion of the ethical safeguards, see (10).]

Three forms of the final examination were constructed. Each form consisted of the same 100 multiple-choice items (four alternative) and differed only in the order of item difficulty: i.e., H-E, E-H, and R. The 100 items were selected from a pool of 197 items administered as a final examination to 931 students enrolled in the same course the previous semester. Item analysis of the 197 questions had yielded item difficulty values (percentage of students passing a given item) for each item. Previous investigators have found that item difficulty values are highly reliable, especially if based on a large group (1, 2, 3, 5, 6). However, since the instructor whose two sections were to be *S*s had taught 107 students of the 931 examined the previous semester, precaution was necessary to avoid teacher contamination of item difficulty values. Accordingly, item difficulties based only on the 107 former students were derived and compared to those based on the entire 931 examined. Twelve items, difficulty levels of which differed significantly (alpha = .05) between the two groups, were eliminated from the pool. From the remaining 185 items, 100 were chosen which had a relatively even spread of item difficulty values ranging from 17.7 per cent to 96.6 per cent and having a mean difficulty level of 62.3 per cent. The items were arranged in three orders: decreasing order of item difficulty, increasing order of item difficulty, and randomly mixed.

The dependent measures were the total number of test items answered correctly and the Multiple Affect Adjective Check List, MAACL (14), which was stapled to the end of each examination. The MAACL measures the amount of anxiety felt in the test situation and was used in this study to examine the effects of the three item difficulty sequences on test-taking anxiety.

A two (anxiety treatments) by three (difficulty orders), unequal *N*, analysis of variance was performed on the achievement test data and on the MAACL scores.

C. RESULTS

Analysis of the achievement test data (total number of items on the final examination answered correctly) revealed no statistically significant differences among the three item difficulty orders ($F < 1.00$, $df = 2/107$, $p > .05$). The other main effect (anxiety treatment) also failed to produce a statistically significant *F* ratio ($F < 1.00$, $df = 1/107$, $p > .05$). Moreover, the interaction of these two variables (anxiety treatments × difficulty orders) resulted in a nonsignificant *F* of 1.38 ($df = 2/107$, $p > .05$).

When the MAACL data were subjected to the same two by three, unequal *N*, analysis of variance, the analysis revealed no statistically significant *F* ratios for either main effects or interaction (alpha = .05). However, a comparison of the mean anxiety score of the 113 *S*s (11.2) to the college norm (6.6) on the MAACL (15) indicates that the *S*s used in this study were extremely anxious at the end of the test.

D. DISCUSSION

So far as this study is concerned, the negative findings with regard to both MAACL and achievement test data obviate assessment of possible mechanisms underlying an item difficulty sequence effect. Such findings are surprising, however, in view of ubiquitous expectations to the contrary. While it may be speculated that negative results on the MAACL data stem from an ineffective anxiety treatment, the Zuckerman and Lubin anxiety norms suggest that the anxiety level inherent in the final examination situation was sufficiently high to mask group differences in anxiety due to treatments. If it can be assumed that test-taking anxiety underlies an existing sequence effect, whatever its form, then a very high initial anxiety would have overshadowed differences in test performance due to the relatively small amounts of anxiety differentially generated by the various item difficulty orders.

Another possible approach to the assessment of sequence effects would be the investigation of individual reactions to test-taking anxiety which would have been masked by the group measures used in this study. Such differential reactions might well interact with item difficulty sequence, and work is progressing at the moment to test this possibility.

In conclusion, this research does not support the hypothesis that there is an item difficulty sequence effect generally operating on the scores of power achievement tests typically found in the classroom. Further, within the limits of this study, it can be concluded that the two major arguments for arranging test items in an easy-to-hard sequence are not valid in any general way.

E. Summary

To investigate item difficulty sequence and anxiety effects on achievement test scores, *S*s were assigned a final examination consisting of items ordered either hard-to-easy, easy-to-hard, or at random, and were tested under conditions of either high or normal anxiety. Data were subjected to a two (anxiety treatments) by three (difficulty orders) analysis of variance which produced no significant *F* ratios, and suggested that individual reactions to test-taking anxiety interacting with item difficulty orders were masked by too general measures.

References

1. Adams, J. F. An evaluation of the effects of level of item difficulty on various indices of item-discrimination. Unpublished doctoral dissertation, State College of Washington, Pullman, Washington, 1959.
2. Brenner, M. H. Test difficulty, reliability, and discrimination as functions of item difficulty order. *J. Appl. Psychol.,* 1964, **48**, 98-100.
3. Carter, H. How reliable are the common measures of difficulty and validity of objective test items? *J. of Psychol.,* 1942, **13**, 31-39.
4. Conrad, H. S. The experimental tryout of test materials. In E. F. Lindquist (Ed.), *Educational Measurement.* Washington, D.C.: Amer. Council on Educ., 1951. Pp. 250-265.
5. Davis, F. B. Item selection techniques. In E. F. Lindquist (Ed.), *Educational Measurement.* Washington, D.C.: Amer. Council on Educ., 1951. Pp. 266-328.
6. Gibbons, C. C. The predictive value of the most valid items of an examination. *J. Educ. Psychol.,* 1940, **31**, 616-621.
7. Greene, H. A., & Jorgensen, A. N. The Use and Interpretation of High School Tests. New York: Longmans, Green, 1939.
8. Jordan, A. M. Measurement in Education. New York: McGraw-Hill, 1953.
9. Lund, K. Test performance as related to order of item difficulty, anxiety, and intelligence. Unpublished Doctoral dissertation, Northwestern University, Evanston, Illinois, 1953.
10. Munz, D. C. The effects of item difficulty sequences on achievement testing scores under two conditions of anxiety. Unpublished Master's thesis, University of Oklahoma, Norman, Oklahoma, 1967.
11. Remmers, H. H., & Gage, N. L. Educational Measurement and Evaluation. New York: Harper, 1943.
12. Ross, C. C. Measurement in Today's Schools (2nd ed.). New York: Prentice-Hall, 1947.
13. Sax, G., & Carr, A. An investigation of response sets on altered parallel forms. *Educ. & Psychol. Meas.,* 1962, **22**, 371-376.
14. Zuckerman, M. The development of an affect adjective check list for the measurement of anxiety. *J. Consult. Psychol.,* 1960, **24**, 457-462.
15. Zuckerman, M., & Lubin, B. Normative data for the multiple affect adjective check list. *Psychol. Rep.,* 1965, **16**, 438.

College of Education
The University of Oklahoma
Norman, Oklahoma 73069

Published as a separate and in *The Journal of Psychology,* 1968, **68,** 185-191.

SHORT TERM MEMORY AND CODING BEHAVIOR AS A FUNCTION OF SET DURING SEQUENTIAL TRIALS*

University of Florida and Kansas State University

E. P. HORNE AND BRUCE C. FLANAGAN

A. INTRODUCTION

Perceptual processing includes sensing, registering, assimilating, and integrating phases. Coding and memory represent an extension of the integrative phase into the responding processes.

Information may be coded on the bases of color, form, or a combination of the properties of these stimuli. Color and form have been considered to be separate discriminations. In early research on similarity matching, Brian and Goodenough (1) indicated that after three years of age color was preferred over form and inferred concepts were based on these cues. Kagan and Lemkin (6) as well as Suchman (7) reported that form was a stronger determinant of conceptual choice than color. Kagan and Lemkin (6) reported that older girls were less likely than younger girls to use color, but older boys used color more than did older girls. The sex difference was attributed to implicit labeling as a basis of similarity judgment. Corah (4) indicated that the amount of color and the complexity of the stimulus forms were related to the number of color responses. Furthermore, there was less response to color when other ways of conceptualizing became more meaningful. From the work of Corah (4) and Gaines (5), certain linkages were noted between perceptual preference and perceptual discrimination.

In Collier's research (2) on visual threshold determination, it was demonstrated that there are a number of associative properties in the measurement. Response linkages were reported to be a function of intertrial intervals. Collier concluded that the probability of a response was a function of luminance and the outcome of preceding trials. Later Collier and Verplanck (3) reported that interpolation, preceding response, and portion of the record, as well as the amount of association, were significant variables. Linkage of response in a single-brightness method was maximized in their experiment.

If linkage is the result of biasing, the response bias was not shown in any

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 4, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

one specific set. The hypothesis of a competing bias, as in a mixed set, as distinctive from a specific, unbiased set either color or form, provides a means of opening the nature of specific phases of bias during the formation of coded responses.

Two experiments were conducted to investigate the processes. In one experiment, the coded response was a B or R: B implied Block or Blue; R implied Round or Red. In the other experiment, the complete word response indicated a coded meaning with the response bias derived from a given set.

B. METHOD

1. *Subjects*

Twenty-four male *S*s for experiment I were obtained from the University of Florida introductory psychology course. The *S*s for experiment II were 24 male volunteers from the beginning behavioral science course at the University of South Florida. Any *S* reporting an uncorrected visual acuity problem or history of color blindness did not participate. It was necessary to replace two *S*s for reasons of reported color blindness.

2. *Apparatus*

The Harvard Tachistoscope was used to present the stimulus material. Exposure time was set at 50 ms with an intertrial interval of 7.5 seconds. Experiment I was carried out in a dimly illuminated 8 × 10 foot room which was isolated from campus noise. Experiment II was conducted in a dimly illuminated 7 × 7 foot sound treated room (IAC 403). In each case, visual barriers blocked *S*s' view of *E*.

3. *Instructions*

In experiment I the following instructions were read to *S*: "In the viewing area you will see a card with two letters on it. Rapidly call out *one* of the two letters immediately following the brief exposure that interrupts this view. If you have any questions the experimenter will be ready to answer them after the completion of the experiment."

In experiment II the *S* received these instructions: "In the viewing area you will see a card with two letters on it. Rapidly call out a *word* that starts with one of the two letters immediately following the brief exposure that interrupts this view. If you have any questions the experimenter will be ready to answer them after completing the experiment."

4. *Procedure*

Each *S* was treated to one of three set stimulus conditions. The color stimulus set consisted of six presentations each of red or blue for the total field. The form stimulus set consisted of six presentations each of round or block-shaped, dark gray forms. The mixed set was three trials each of (*a*) a red round form, (*b*) a blue round form, (*c*) a red block, and (*d*) a blue block. All forms were approximately one inch in diameter and centered on a light gray exposure field. Eight of the *S*s in each experiment were treated to one of the above stimulus set conditions. The sequence of presentations within the set was ordered by a random schedule restricting recurrence of like conditions to two successive trials.

Pre- and poststimulus fields contained two letters R and B, positioned vertically. Odd numbered *S*s were run with B in the top position, and even numbered *S*s with the reverse letter position.

In trial sequences which followed set conditions, there were 48 presentations with four basic orders: ABCD, BDAC, CADB, and DCBA. The A order of stimulus cards was blue-round, red-block, red-block, blue-round, red-round, red-round, blue-block, blue-round, blue-block, red-round, red-block, blue-block. The B order was red-block, blue-block, red-round, red-round, blue-block, red-block, blue-round, blue-round, red-block, blue-block, blue-round, red-round. The C order was blue-round, red-round, red-block, blue-block, blue-round, red-block, red-round, blue-round, blue-block, red-round, red-block, blue-block. The D order was red-block, blue-block, red-block, red-round, blue-round, blue-round, red-round, blue-block, red-round, blue-round, blue-block, red-round. The 24 *S*s of each experiment were used with six *S*s in each of the four orders.

Experimental sessions progressed from set condition to trial sequence without informing *S* and with no change in intertrial time.

5. *Scoring*

For the coded experiment in which *S* was instructed to call out either R or B after each stimulus presentation, the presence of a B response was marked by *E* on a score sheet by trial. In experiment II, in which a word response was required, *E* recorded the word called out for each trial. Data were scored according to the first letter, yielding similar data to experiment I. Words were also scored according to their relevance to the stimulus presented as relevant color, relevant form, or irrelevant.

C. RESULTS

1. *Experiment I*

Where the coded R or B responses were required, the data were treated to three-dimensional analysis of variance regarding set, sequence, and stimuli as main effects. The results of this analysis are presented in Table 1. All main effects—set, sequence, and stimulus properties—were significant, and the interactions were not. Therefore when a letter, which had ambiguous (nonmeaningful) but coded value, was elicited, the exposure variable was generally effective. The 50 ms exposure (mixed set-color set-form set) which was designed to establish a specific set was significant at the .05 level. The *t* tests for color set and form set were both significant (.05) when compared with mixed set conditions, but color effect *vs.* form set were not significantly different from one another.

TABLE 1
CODED RESPONSE B

Source	*df*	SS	MS	*F*
A (set)	2	50.02	25.01	3.19*
B (sequence)	3	118.75	39.58	5.05**
C (stimuli)	3	560.66	186.89	23.84**
A × B	11	169.06	15.37	1.96
A × C	11	162.40	14.76	1.88
B × C	15	151.59	10.11	1.29
Error	50	391.85	7.84	
Total	95			

* .05 level.
** .01 level.

The sequence variable, arrangement of red-block, blue-block, red-round, blue-round introduced as fixed orders to insure among the 24 *S*s a balanced arrangement of these four cards, was found to be significant (.01) level. Of the six *t* tests for sequence, three were significant at .05 and one was significant at .01 level. The last significant *t* was between the two orders, one inverted from the other.

The stimulus properties—red-block, blue-block, red-round, and blue-round attributes—were significant in the determination of the coded response. In six *t* tests, four differences—red block-blue block, red block-blue round, blue block-red round, and red round-blue round—were significant at .01 level. Where a forced choice of one response excludes an alternative response, color set was more effective in the emitted coded response.

In terms of coded responses, both color set and form set were significantly different from the mixed set in determining the response. There was no significant difference between color set and form set.

Furthermore, the sequences for test pairs of color-form, which were arranged in four balanced ordered sequences, were significant.

When *S*s were responding on a short-term memory function with only the coded response of one letter (the initial one of a possible word), the experimental variables of set and sequence were significant in determining that coded response. Although the code was not interpreted to the *S*, the implication was present, although the particular aspect of code—color *vs.* form—was not stated by direction or by explicit meaning.

2. *Experiment II*

In a second experiment with 24 other *S*s most of the procedure was replicated, but the instruction to the *S* was altered. The second group was asked to respond by giving a word.

TABLE 2

WORD RESPONSE B

Source	*df*	SS	MS	*F*
A (set)	2	7.77	3.88	—
B (sequence)	3	36.71	12.24	—
C (stimuli)	3	219.71	73.24	5.38*
A × B	11	15.23	1.38	—
A × C	11	114.73	10.43	—
B × C	15	41.56	2.77	—
Error	50	680.27	13.60	
Total	95			

* 3/50 *df*, 4.20 = .01 level.

An analysis of variance of word responses (Table 2) scored for initial letter B (B-words) indicated that the stimulus properties—red-block, blue-block, red-round, blue-round attributes—were significant in the determination of the B-word responses. Of the six *t* tests, three were significant at the .01 level and one was significant at the .05 level. These four differences were red block-blue block, red block-blue round, blue block-red round, and red round-blue round. It may be noted that the pattern of response in the two experiments with coded (letter) response and with word response was similar. In both experiments BB (blue block) gave highest *B* response or *B-word* response. In both experiments the RR (red round) gave low *B* response.

The color code was apparently producing more names in completing a B response when two cues were not implicitly consistent: i.e., block-red or blue-round. Coding and meaning were either ambiguous or in conflict with only partial cues. Since there were no instructions given to the subjects that letter was a definite lead to a word response, the subjects were able to use a word (nonrestrictive to color or form). The low level of random naming indicated that behavior was particularly selective with color cues stronger than form cues.

Immediate memory response was influenced by selective cues. The significance level of the stimulus effect was .01.

In an attempt to interpret the responses, three further analyses of variance were done: scoring for (*a*) relevant color response, (*b*) relevant form response, and (*c*) irrelevant word response. It was noted that the main effect of set was significant for irrelevant words, relevant color responses, but not for relevant form. Furthermore, sequence and stimuli, when scored for relevancy, were not significant. In two of the three analyses, the set stimulus interaction was significant but not the form response.

From the data of experiment II, the presence of a few cues was noted—an additional interpretation of the character of the responses: irrelevant word, relevant color, and relevant form, the low effect of form as a particular cue in terms of significance. The stimulus-set interaction just compounded the simple effect. Perceptual responses were not completely reduced to simple effects.

D. INTERPRETATION

Coding by letter was determined by set. Color, as opposed to form, as a separate discriminatory effect was not more effective for the separate letter response. Within the experimental set for limited trials, the B or R response has not been biased by color or form. The stimulus information apparently was inadequate, whereas the set information was effective.

The significant sequence effect indicated that subjects were coding on a nonrandom basis. The memory and coding behavior may have been a form of response bias as in the Collier (2) and the Collier and Verplanck (3) experiments. Stimulus information did not register. One implication which is suggested is that registration-assimilation processes were set linked.

As the information in the stimulus properties for the mixed and unmixed sets increased to two levels—red-block, blue-block, red-round, and blue-round—significant effects were obtained. In coding with a separate letter response, the greater stimulus input prepared the subject for registration and interpretation.

In the second experiment the word response produced more coding but with only a limited meaning: ambiguous or irrelevant responses. Since random names appeared at a low frequency, the behavior had been selective. Sequence, when scored for relevancy, was not significant.

The form-set effect may not be attributed to difficult form discrimination but probably to some difficulty in producing meaning within the coding behavior.

E. CONCLUSION

Perceptual processes as determiners of response were informationally limited through color and form properties. Where the informational content was increased, a code effect was observed. Meaning, as it is limited to a word response, was not completely clear to these subjects, probably in relation to incomplete instruction. Associative behavior is complex in linkage.

F. SUMMARY

In mixed set with competing bias, as compared with a specific, unbiased set, perceptual processes were informationally limited through color and form properties. With one instruction in the first experiment, responses were R or B. The second experiment required a word response. Selective cues regulated alphabetical choice, which probably indicated coding behavior. In the formation of a word response, coding was limited by ambiguous or irrelevant responses. Specific response linkage was not generally obtained.

REFERENCES

1. BRIAN, C. R., & GOODENOUGH, F. L. The relative potency of color and form perception at different ages. *J. Exper. Psychol.*, 1929, **12**, 197-213.
2. COLLIER, G. Intertrial association at the visual threshold as a function of intertrial interval. *J. Exper. Psychol.*, 1954, **48**, 330-334.
3. COLLIER, G., & VERPLANCK, W. S. Nonindependence of successive responses at the visual threshold as a function of interpolated stimuli. *J. Exper. Psychol.*, 1958, **55**, 429-437.
4. CORAH, N. L. The influence of some stimulus characteristics on color and form perception in nursery-school children. *Child Devel.*, 1966, **37**, 205-211.
5. GAINES, R. Color-form preferences and color-form discriminative ability of deaf and hearing children. *Percept. & Motor Skills*, 1964, **18**(1), 70.
6. KAGAN, J., & LEMKIN, S. Form, color, and size in children's conceptual behavior. *Child Devel.*, 1961, **32**, 25-28.
7. SUCHMAN, R. G. Color-form preference, discrimination accuracy and learning of deaf and hearing children. *Child Devel.*, 1966, **37**, 439-451.

Department of Psychology
University of Florida
Gainesville, Florida 32601

Published as a separate and in *The Journal of Psychology,* 1968, **68,** 193-201.

RESULTS OF THE NATIONAL TEACHER EXAMINATIONS AT TULANE UNIVERSITY, 1948-1966*

Tulane University

E. C. Hunter

The National Teacher Examinations have been administered to teacher education seniors at Tulane University annually from 1948 to the present. The examinations provide measures for evaluating academic preparation in the three main areas of preservice teacher education: general education, professional education, and teaching area specialization. These areas comprise the chief components of teacher education programs in most colleges and universities in the United States.

Developed and conducted by Education Testing Service since 1950 and by the American Council on Education from 1940 to 1950, the National Teacher Examinations are used by many local school systems as an aid in selecting and promoting teachers, by certain State Departments of Education for certification purposes, and by teacher education institutions for evaluating achievement of their students and effectiveness of their teacher education programs. The Common Examinations have generally included tests of Professional Information, Written English Expression, Social Studies-Literature-Fine Arts, Science and Mathematics, and Nonverbal Reasoning. The last-named test was discontinued after 1964. Teaching Area Examinations have been available in some 13 fields, including two in elementary education, Biology and General Science, English Language and Literature, Industrial Arts Education, Mathematics, Chemistry-Physics-General Science, Social Studies, Physical Education, Business Education, Music Education, Home Economics Education, and Art Education. Total testing time is approximately six hours.

The examinations are designed for college seniors who have substantially completed Bachelors degree programs of preservice teacher preparation. The tests do not claim to encompass all factors in teacher preparation, such as motivation, attitudes, personality, and values. They do cover cognitive factors and are designed to assess knowledge, understanding, and mastery of subject matter in specialized areas.

Scores on separate tests, as well as weighted totals, are reported in terms of

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 4, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

Published as a separate and in *The Journal of Psychology*, 1968, 68, 193-201.

RESULTS OF THE NATIONAL TEACHER EXAMINATIONS AT TULANE UNIVERSITY, 1948-1966*

Tulane University

E. C. HUNTER

The National Teacher Examinations have been administered to teacher education seniors at Tulane University annually from 1948 to the present. The examinations provide measures for evaluating academic preparation in the three main areas of preservice teacher education: general education, professional education, and teaching area specialization. These areas comprise the chief components of teacher education programs in most colleges and universities in the United States.

Developed and conducted by Education Testing Service since 1950 and by the American Council on Education from 1940 to 1950, the National Teacher Examinations are used by many local school systems as an aid in selecting and promoting teachers, by certain State Departments of Education for certification purposes, and by teacher education institutions for evaluating achievement of their students and effectiveness of their teacher education programs. The Common Examinations have generally included tests of Professional Information, Written English Expression, Social Studies-Literature-Fine Arts, Science and Mathematics, and Nonverbal Reasoning. The last-named test was discontinued after 1964. Teaching Area Examinations have been available in some 13 fields, including two in elementary education, Biology and General Science, English Language and Literature, Industrial Arts Education, Mathematics, Chemistry-Physics-General Science, Social Studies, Physical Education, Business Education, Music Education, Home Economics Education, and Art Education. Total testing time is approximately six hours.

The examinations are designed for college seniors who have substantially completed Bachelors degree programs of preservice teacher preparation. The tests do not claim to encompass all factors in teacher preparation, such as motivation, attitudes, personality, and values. They do cover cognitive factors and are designed to assess knowledge, understanding, and mastery of subject matter in specialized areas.

Scores on separate tests, as well as weighted totals, are reported in terms of

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 4, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

"Scaled Scores." These are standard scores with a mean of 60 for National Teacher Examinations candidates which permit direct comparisons among tests in a given year and from year to year.

Local scaled scores may be compared section by section or with various norms groups. Norms data are based on scores made by a sample of National Teacher Examinations candidates from all parts of the country annually. The number of candidates comprising the sample from which the norms data are derived runs to seven or eight thousand. Since the sample of institutions which test nearly all of their full-time teacher education seniors is carefully selected on a stratified basis, this number of seven or eight thousand students comprising the norms group is adequate statistically. Comparisons and interpretations in this study will be made in terms of percentile ranks of scaled scores. For instance, a scaled score of 62 on the social studies test has a percentile rank of 70, which means that on this test the score of 62 ranks 70th from the bottom or 30th from the top in a group of 100 candidates whose scores are distributed exactly as the several thousand scores in the national sample were distributed. School systems, of course, may develop their own norms so that an individual's scaled score may have different percentile rank values in different school systems. Interpretations here are limited to comparisons with the national norms provided by Educational Testing Service. Percentile ranks are based on candidates tested over the country during 1957-1961 from a stratified sample of institutions which tested almost all full-time seniors preparing to teach. This sample was selected for comparison with Tulane results because 1957 marked the middle of the 19-year period under study.[1]

Reliabilities of the Common Examinations tests range from .80 to .90, and for the Professional Education and General Education Weighted Subtotal Scores and the Weighted Common Total Scores from .93 to .96. Concurrent validity is indicated by the relationship between scores on the National Teacher Examinations and the criterion of grade-point average. For seniors enrolled in teacher education programs at 11 institutions, the correlations range from .38 to .74 with an average of .57. Content validity is indicated by the care taken by subject matter committees and test specialists in writing and selecting items. Following each administration, rigorous statistical analyses of the items are made to assure that the examinations meet high standards.[2]

Table 1 shows the number of Tulane students completing Bachelor's and

---

[1] National Teacher Examinations, *Interpretation of Scores, K2.* Educational Testing Service, Princeton, New Jersey, 1957-1961.

[2] National Teacher Examinations, *Technical Handbook.* Educational Testing Service, Princeton, New Jersey, 1965. Pp. 13-18.

Master's degree and teacher education programs together with number and percentage taking the National Teacher Examinations by years from 1948 to 1966. Of 1415 students completing undergraduate and graduate programs in teacher education over the 19 years, 693 or 49.1 per cent took the National Teacher Examinations. Percentages ranged from 14.7 in 1965 to 96.0 in 1957. Students have been requested but not required to take the tests. After 1952 the number taking the tests increased because the New Orleans Public Schools began requiring the scores for regular employment. During the past two years of the study, quite small percentages of those completing programs took the Examinations. Included in Table 1 are the scores of 25 of 270 graduate students who did not take the tests when they were graduating seniors, but took the tests during the pursuance of their Master's program in education. The 693 or 49.1 per cent of all students completing programs represents an adequate Tulane University sample for group interpretations on the National Teacher Examinations. This number comprised approximately 90 per cent of all students who actually took teaching jobs after graduation.

TABLE 1

NUMBER OF TULANE UNIVERSITY STUDENTS COMPLETING BACHELOR'S AND MASTER'S DEGREE AND TEACHER EDUCATION PROGRAMS TOGETHER WITH NUMBER AND PERCENTAGE TAKING NATIONAL TEACHER EXAMINATIONS BY YEARS FROM 1948 TO 1966

Year	Number completing programs	Number taking National Teacher Examinations	Per cent taking National Teacher Examinations
1948	45	20	44.4
1949	62	24	38.7
1950	78	27	34.8
1951	99	18	18.1
1952	78	40	51.2
1953	63	24	38.1
1954	53	29	54.5
1955	46	37	80.4
1956	51	38	74.5
1957	51	49	96.0
1958	79	50	63.3
1959	92	53	57.6
1960	118	74	62.8
1961	100	40	40.0
1962	105	54	51.4
1963	93	37	39.7
1964	80	57	71.2
1965	61	9	14.7
1966	61	13	21.3
Total	1415	693	49.1

Composite results on the National Teacher Examinations at Tulane University will be set forth showing scaled scores, means, medians, and percentile ranks for the 19-year period.

Table 2 shows the distribution, mean, median, and percentile rank of the Weighted Commons National Teacher Examinations total scores for Tulane University teacher education students over the past 19 years. The median of 627 and percentile rank of 68 indicate that the selection and training of teacher education students at Tulane University have been quite above the norms for the 19-year period. Actually the median range was 135 points, 540 in 1949 for 24 students to 675 in 1966 for 10 students. The median for the last five years combined, 1962 to 1966, was 654 for 152 students. This gain in median Weighted Commons total score of 61 points from 1951-1955 to 1962-1966 was significant at beyond the .01 level. Thus, it can be said with high confidence that the selection and training of teacher education students at Tulane University improved substantially over the past 15 years.

TABLE 2
DISTRIBUTION, MEAN, MEDIAN, AND PERCENTILE RANK OF WEIGHTED NATIONAL TEACHER EXAMINATIONS COMMONS TOTAL SCORES OF COMBINED TEACHER EDUCATION STUDENTS COMPLETING BACHELOR'S AND MASTER'S PROGRAMS AT TULANE UNIVERSITY 1948 TO 1966

Score	Total number students
800	7
750	21
700	78
650	158
600	182
550	128
500	78
450	25
400	14
350	1
300	1
N	693
Mean	624
Median	627
Percentile rank	68

Table 3 shows distributions, means, medians, and percentile ranks of teacher education students at Tulane University on the various sections of the Commons Examinations of the National Teacher Examinations for 19 years, 1948 to 1966.

On the Professional Information section, comprising Psychological Foundations, Societal Foundations, and Teaching Principles and Practices, the median scaled score for 693 Tulane teacher education students over 19 years was 65, which was associated with a percentile rank of 73. This means that only 27 per cent of the national sample made scaled scores above 65. One hundred seventy-nine, or 26 per cent, of the Tulane students made scaled scores of 70 or above with percentile ranks of 87 or above. Forty-two per cent made scaled scores between 60 and 70, while only 32 per cent had scores below 60, the average achieved by National Teacher Examinations candidates. The superiority of the Tulane teacher education students on the Professional Information section of the national test is clearly apparent. The mean scaled score for 187 students during the first five years of the 19-year period on Professional Information was 59; that for 152 students during the last five years was 66. This difference of seven points was significant at beyond the .01 level. Thus, the selection and training of Tulane students as judged by the Professional Information test improved substantially from the early 1950's to the middle 1960's.

TABLE 3

DISTRIBUTIONS, MEANS, MEDIANS, AND PERCENTILE RANKS OF SCALED SCORES OF TEACHER EDUCATION STUDENTS AT TULANE UNIVERSITY ON PROFESSIONAL INFORMATION, WRITTEN ENGLISH EXPRESSION, SOCIAL STUDIES-LITERATURE-FINE ARTS, NONVERBAL REASONING, AND SCIENCE AND MATHEMATICS SECTIONS OF THE NATIONAL TEACHER EXAMINATIONS FOR 19 YEARS, 1948-1966

Scaled score	Professional information	Written English Expression	Social Studies-Literature-Fine Arts	Science and Math.	Nonverbal Reasoning
95				2	
90			1	2	
85	4		2	9	1
80	12	8	19	16	7
75	47	46	46	36	43
70	116	125	75	80	85
65	170	151	104	118	151
60	118	134	108	147	177
55	101	84	141	134	96
50	56	70	115	84	48
45	43	42	60	44	17
40	16	20	16	15	18
35	8	8	5	6	10
30	2	5	1		3
N	693	693	693	693	656
Mean	63	63	61	62	63
Median	65	65	60	62	63
Percentile rank	73	70	62	64	62

Table 3 shows that, on the Written English Expression section of the National Teacher Examinations, 693 Tulane students over 19 years achieved a median scaled score of 65 which was equivalent to a percentile rank of 70.

Only 30 per cent of the national sample made scaled scores as good as or above 65 on this test. One hundred seventy-nine, or 26 per cent, made scaled scores of 70 or above with percentile ranks of 89 or above. Forty-one per cent of the Tulane students made scaled scores on the English test between 60 and 70. Thus, only 229, or 33 per cent, of the Tulane teacher education students made scaled scores on the English test below 60, the average for National Teacher Examinations candidates. The mean scaled score for 187 students during the first five years was 58; that for 152 students during the last five years was 66. This mean difference of eight points was significant at beyond the .01 level. Thus, 67 per cent of the Tulane students on the English Expression Test over 19 years exceeded the mean of the national sample. Also, the selection and training of the Tulane students as judged by the English test improved appreciably from the early 1950's to the middle 1960's.

On the Social Studies-Literature-Fine Arts section of the National Teacher Examinations, the scaled score median for 693 Tulane teacher education students over the 19 years was 60, which had a percentile rank of 62. On this test 38 per cent of the national sample made scores as good as or above 60. One hundred forty-three, or 21 per cent, of the Tulane students made scaled scores of 70 or above, falling in the upper 9 per cent of the national sample. Thirty-one per cent made scores between 60 and 70, while 338 or 48 per cent had scaled scores below 60. Thus, the Tulane teacher education students over the 19 years ranked just above the national sample on the Social Studies test. The mean scaled score for 187 students during the first five years of the study was 58; that for 152 students during the last five years was 64. This mean difference of six points was significant at the .01 level. Hence, the selection and training of Tulane teacher education students as judged by the Social Studies-Literature-Fine Arts section of the National Teacher Examination showed definite improvement from the early 1950's to the middle 1960's.

Table 3 further shows that on the Science and Mathematics section of the Commons Examination, the median scaled score for 693 Tulane teacher education students over 19 years was 62, with a percentile rank of 64. Twenty-one per cent made scores of 70 or above, falling in the upper 15 per cent of the national sample. Thirty-eight per cent had scaled scores between 60 and 70, while 41 per cent made scaled scores on this test below 60. It appears that the Tulane students, on the average, over the 19 years made appreciably higher scaled scores on the Science and Mathematics general education test than did the national sample. The mean scaled score for 187 students during the first five years of the study on this section was 60; that for 152 students during the last five years was 65. This mean difference of five scale score points was significant at the .01 level. Thus, as judged by the Science and Mathematics test, the selection and training of teacher education students at Tulane University improved appreciably from the early 1950's to the middle 1960's.

On the Nonverbal Reasoning section of the Commons Examinations, discontinued after 1964, the median scaled score for 656 Tulane teacher education students was 63 with a percentile rank of 62 on the national sample. One hundred thirty-six, or 21 per cent, had scaled scores of 70 or above, falling in the upper 14 per cent of the national sample. Fifty per cent of the Tulane students made scaled scores between 60 and 70, while 29 per cent had scores below 60, the average of the national sample. Thus, the selection of Tulane teacher education students, typically, from 1948 to 1964, as judged by the Nonverbal Reasoning test, was slightly better than that of the National Teacher Examinations sample. The mean scaled score for 187 students during the first five years of the study was 61; that for 219 students during the last five years was 65. This mean difference of four points was significant at the .01 level. Thus, as judged by the Nonverbal Reasoning test, the selection of teacher education students at Tulane improved appreciably from the early 1950's to the early 1960's.

TABLE 4

DISTRIBUTION, MEAN, MEDIAN, AND PERCENTILE RANK OF SCALED SCORES OF TEACHER EDUCATION STUDENTS AT TULANE UNIVERSITY ON COMBINED NATIONAL TEACHER EXAMINATIONS OPTIONAL OR TEACHING AREA TESTS FOR 19 YEARS, 1948 TO 1966

Scaled score	Number of students
80	13
75	30
70	87
65	135
60	160
55	94
50	48
45	21
40	8
N	596
Mean	63
Median	64
Percentile rank	65

Table 4 presents the results of 596 Tulane University teacher education students over 19 years on the combined Teaching Area tests of the National Teacher Examinations. The median scaled score was 64 with a percentile rank of 65. One hundred thirty, or 22 per cent, of the students made scaled scores above 70, or in the upper 16 per cent of the national sample. Two hundred ninety-five, or 50 per cent, of the students had scaled scores between

60 and 70, while 28 per cent made scores below 60, the average of the national sample. These data indicate that the Tulane teacher education students, typically over the 19 years on the Teaching Area tests, were substantially better than the average for the country. The mean scaled score for 127 students during the first five years of the study was 62; that for 149 students during the last five years was 65. This mean difference of three points was significant at the .05 level. Thus, there is evidence that the selection and training of teacher education students at Tulane University as judged by the Teaching Area tests improved slightly from the early 1950's to the middle 1960's.

On the combined Teaching Area tests, Tulane University teacher education students showed a significant gain in mean scaled scores from the first five to the last five years of the 19-year period. When results were analyzed for the tests separately, two of the Area tests, Education in the Elementary School and English Language and Literature, showed no significant change in means from the first five to the last five years of the study.

Significant improvement in selection and preparation of teacher education students at Tulane University over 19 years, 1948 to 1966, is clearly indicated in the results of the National Teacher Examinations presented herewith. This is true for all areas tested, general education, professional education, and teaching area specializations. Although the general selective admissions policy at Tulane has not been materially changed during the period of this study, the quality of students admitted to the University and to teacher education programs has been improved. Also, it seems reasonable to conclude that some of the increase in scores from the first five years to the last five years of the 19-year period can be attributed to improved teaching in certain subject matter departments and in professional education.

SUMMARY

The purpose of the study was (*a*) to compare Tulane University Teacher Education students with national norms on the National Teacher Examinations over the past 19 years, and (*b*) to investigate the selection and preparation of the Tulane students during the first five years and the last five years of the 19-year period. The data indicated that the Tulane teacher education students exceeded the norms every year since the tests were initiated in 1948. Further, the data showed significant improvement in the selection and preparation of students from the first to the last years of the period. This was true in all areas tested: general education, professional education, and teaching area specializations. Since Tulane University's selective admissions policy has not been mate-

rially changed during the period of the study, it seems reasonable to attribute some of the gains in scores from the first to the last years of the period to improved teaching in subject matter departments and in professional education.

Center for Teacher Education
Tulane University
New Orleans, Louisiana 70118

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 203-209.

COLOR RESPONSE TO THE TAT: AN INSTANCE OF EIDETIC IMAGERY?*[1]

The City College of The City University of New York

PETER W. SHEEHAN

A. INTRODUCTION

In the course of a study on posthypnotic conflict (16), a subject was put into trance, awakened, and given a battery of personality tests to determine base control measures of personality functioning. This procedure was followed prior to the induction of an anxiety-toned conflict. Asked by *E* to make up a story to card 3 BM of the Thematic Apperception Test (TAT) (9), *S* reported the following: "I can see some colors in the picture. This sounds a bit looney, but would you like me to tell you about the colors? Her hair is a dark brown; she has a swarthy complexion. Her skirt is a light blue. Her sweater is a reddish brown color; it's dirty. The carpet is green and the sofa is yellow." When asked the following day to make up a story to the card, *S* reported colors again but attributed different colors to the same parts of the stimulus.

To the author's knowledge there has been no other instance of color response to the TAT reported in the literature. When color is projected onto an achromatic area of the Rorschach cards (12), "the fantasy dominated quality of the response suggests a serious loosening of ties with reality and an inappropriate manifestation of affect" (8, p. 339). The possible pathology of color response to the TAT is highlighted by the fact that inappropriate color responses in the Rorschach may be suggested by the association of color with the achromatic features of cards 2 and 3. The TAT in fact is totally achromatic and offers no such cue-eliciting suggestions for color responses.

Color responses to achromatic stimuli in projective tests undoubtedly reflect some degree of pathology in many instances. But an alternative hypothesis for explaining this behavior which should at least be investigated is that *S*s may

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 7, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was carried out while the author was working as research associate at the Unit for Experimental Psychiatry, the Institute of the Pennsylvania Hospital and the University of Pennsylvania. It was supported in part by Grant No. NONR 4731(00) from the Office of Naval Research.

have an unusual propensity for vivid imagery of the eidetic kind: they may project colors onto stimuli in vivid eidetic images which persist through time. Such behavior emphasizes less their personality functioning than the cognitive factors which they bring to the task. Accordingly, the present study was designed to investigate extensively this particular *S*'s imagery and the conditions of its occurrence. Only close study of *S*'s imagery can determine whether it is unique enough to account for his color responses to the TAT.

B. Method

The *S* was a 21-year-old, male university student. Prior to imagery testing, several relevant tests were administered. A high score of 11 on the Stanford Hypnotic Susceptibility Scale, Form C (18), and a maximum rating of 5 for trance depth on a diagnostic scale of hypnotic depth (11) indicated that he was very susceptible to hypnosis. He scored an *IQ* of 120 on the Quick Test of intelligence (1) and showed a normal adjustment score on the Incomplete Sentences Blank, College Form (13). His score of 147 on the sentence completion test was +.45 standard deviations away from the mean adjustment score given by 14 *S*s of similar hypnotic susceptibility. An electroencephalographic record was also taken and examined for clinical indications of epilepsy. No evidence was found of any disturbance in brain wave patterns.

S was tested for eidetic imagery in a standarized manner adopted from the techniques of Haber and Haber (5). He was instructed first to scan the stimulus and then look hard at a projected screen placed on an easel and to tell *E* what he saw. Instructions to scan ensured that afterimages were not formed following fixation of the stimulus figure. Seven stimuli were used. *S* was shown three geometrical designs, card 3 BM of the TAT, a colored stimulus depicting an aeroplane crash, and two stimuli from Witkin's embedded figures test (19). He was asked to construct a copy of his image of each geometrical pattern by means of blocks, to describe his images of the TAT card and plane crash verbally, and to draw copies of his images of the embedded figures. Verbal description was not used to test imagery of the designs, since previous evidence has shown that reconstruction is a more sensitive technique than verbal recall for measuring the accuracy of *S*s' imagery of simple stimuli (14). After *S* had imaged the TAT card and plane crash, he was shown both stimuli and asked to describe them, while *E* checked on the accuracy of his perception. Each image was rated by *S* on the Betts seven-point rating scale for vividness of imagery (2).

Following imaging of the experimental stimuli, *S* completed a shortened form of the Betts' Questionnaire Upon Mental Imagery (15).[2] A general inquiry was then conducted into *S*'s use of different sensory modalities in his imagery, the nature of his dream imagery, the pleasantness and familiarity of the stimuli recalled, and the frequency of spontaneously occurring eidetic imagery outside the experimental setting. Following this inquiry he was tested for negative afterimages to the imaged colors: red, blue, yellow, and green. Here he was asked to image a color for 30 seconds while fixating a center dot on a white projection field, and then to transfer his gaze to a grey projection field and report what he saw. He was then tested for negative afterimages to the actual colored stimuli. Finally, he was tested for defective color vision on the AO H-R-R pseudochromatic plates (6) under standard conditions of illumination.

C. Results

Following the investigations of Haber and Haber (5) and Siipola and Hayden (17), five primary criteria were adopted for distinguishing the eidetic image from the afterimage and the memory image. The criteria were as follows: *E* noted whether (*a*) an image was reported, (*b*) the color was positive, (*c*) an image was located "out-there" or in *S*'s mind, (*d*) *S* used the present or past tense in his verbal descriptions, and (*e*) he moved his eyes around the projection field in the proper relative directions when describing different areas of the image.

Data indicated that *S* had an unusual propensity for vivid imagery. Comparison of his performance with the average rating performance of a large sample of college students on whom imagery test data were available (15) showed that *S* imaged more vividly on the imagery questionnaire in all sensory modalities. Ratings differed most appreciably in the visual, auditory, gustatory, and kinaesthetic modalities. Considering all items, *S* scored an average rating of 2.0 ("very clear and vivid") on the seven-point scale of vividness of imagery.

S's behavior showed strong evidence of eidetic imagery. Four of the seven images which *S* evoked satisfied the criteria for eidetic imagery as stipulated by Haber and Haber (5). Eidetic images were given to the TAT card, the scene of the aeroplane crash, and to both of the embedded figures. All four images were rated as either moderately clear and vivid or very clear and comparable in vividness to the actual experience, were positively colored, and were seen as "out-there" in the projection field. The most vivid and stable image was of the aeroplane crash which *S* described in a 665-word description given entirely in the present tense. He made only three errors in the recall of this stimulus, all of which were concerned with relatively detailed parts of the picture. Other details were recalled with remarkable accuracy even down to the symbol on the tie pin of one of the figures depicted in the picture. However, accuracy of reproduction was not perfect for the other stimuli, which is in accord with the findings of Doob (3) and Siipola and Hayden (17) that eidetic imagery may not be highly correlated with accuracy of recall, at least in some instances. Images of the embedded figures and the TAT card were not consistently stable and somewhat less vivid than *S*'s image of the plane crash, but in all four eidetic images *S* experienced no difficulty in pointing to different parts of the projection field and describing corresponding sections of his images. When *E* tested *S*'s imagery of one of the embedded figures, he pointed to the screen and asked *S* what was there. The *S* replied that he could see nothing at the spot *E* was pointing to, but informed *E* that he had gone past a 2″ by 1″ image situated directly in the middle of the projection screen. It was significant that the TAT card which was imaged eidetically was again reported as colored both in *S*'s image and in his description of the actual card.

---

[2] The scale contains 35 items of the original Betts questionnaire (2). Five items pertain to each of seven sensory modalities; they are visual, auditory, cutaneous, kinaesthetic, gustatory, olfactory, and organic. In the visual (auditory) modality, *S*s are for example asked to think of seeing (hearing) "the sun as it is sinking below the horizon" ("the sound of escaping steam") and to consider carefully the image which comes to their mind's eye (ear).

The three geometrical designs reproduced by blocks were not imaged eidetically. No image of the designs satisfied the criteria laid down to detect the presence of eidetic imagery. None was projected externally, any verbal report on the imagery was in the past tense, and the imagery was reported typically as not clear or vivid.

An important feature of *S*'s behavior was the sporadic nature of the imagery. Not all stimuli were imaged eidetically. *S* reported in the inquiry that he could experience vivid auditory imagery of an eidetic kind: "If you mention to me the first movement of Brahms Fourth, I can actually hear it and I can hear the strings in perfect detail and if I can concentrate long enough and you gave me some staff paper I could perhaps pick out the second viola part and write it down for you . . . sometimes I get it (image of Brahms) and sometimes I don't, but when I get it I usually get it pretty well. You just forget things sometimes." The factors which appeared to influence the occurrence of eidetic imagery were *S*'s familiarity with the imaged stimulus and the affect associated with it. Scenes which were familiar and had strong affect attached to them were imaged vividly. *S* reported the greatest intensity of imagery in the visual modality. Visual images had "strength and vividness above auditory experiences because of the emotional intensity associated with them."

In the inquiry *S* reported that he sometimes saw black and white movies in color. Paradoxically, this phenomenon occurred usually when the actresses, actors, and scenery in the movie were not very familiar. He reported color as not actually on the screen, but in his mind. Doubt, however, about the external quality of his imagery reflects more *S*'s belief in the unreality of the image. His verbal descriptions vividly depicted an external, "out-there" quality characteristic of eidetic imagery. The TAT card was again dsecribed as a "water color" and *S* reported that nine-tenths of his dreams were always in vivid color. These data highlight the peculiar emphasis *S* gives to color. When tested on the color blindness test, *S* showed perfectly normal color vision.

S reported normal afterimages to actual colors but failed to show unequivocal evidence of negative afterimages to imagined stimuli. On three of the four test occasions when he fixated the projection field after imaging a colored stimulus, *S* continued to project the same colored image he had been initially asked to evoke. On the discrepant trial, after imaging the yellow stimulus, he transferred his gaze to the projection field and failed to see anything. But on the next trial *S* saw both a yellow and a blue square after imaging the latter. It seems more appropriate to regard the yellow image on this trial as a projection of the previous image than a negative afterimage to the blue stimulus, particularly since juxtaposition of two eidetic images on the same projection field is a phenomenon which is known to occur in the imagery of some eidetic subjects (10).

D. Discussion

The results suggest that the color responses to the TAT initially observed in this *S* can more plausibly be explained by the hypothesis that he has a special aptitude for eidetic imagery than by appealing to hypotheses concerning psychopathology of adjustment. He showed no disturbance on the Rotter sentence completion test and no unusual responses to the TAT other than color naming. He was capable of exceptionally vivid imagery and showed evidence of sporadic eidetic ability. Factors which determine the strength and vividness of imagery were familiarity with the stimulus, the affect attached to it, and the type of sensory modality employed. For this *S*, vivid eidetic imagery favored familiar material though color was easily projected onto unfamiliar objects. The presence of at least some affect was the more important determinant of eidetic imagery. Vivid eidetic images occurred more often in the visual and auditory modalities than in other modalities because of the stronger affect attached to them.

Motivation and interest played a special role in his eidetic imagery. *S* appeared not to project images externally which had little affect attached to them. This is consistent with previous reports that eidetic imagery can occur for some kinds of pictures but not for others (4). Eidetic imagery seems to reflect a cognitive ability which has survived from some earlier evolutionary state and which has become virtually functionless in adult society (4). The survival of such imagery in some *S*s suggests it is performing a useful function. Perhaps as Doob suggests it is activated when the individual is experiencing some special kind of difficulty in coping with the environment (4). The reason for its survival and the role motivation and interest play constitute the true clinical relevance of its occurrence.

It is important to consider whether the occurrence of eidetic imagery in the present context was genuine or merely a function of *E*'s implicit suggestions concerning its occurrence. Instructions to report what *S* sees on a projected field certainly imply that eidetic imagery is being investigated. The criticism is especially pertinent considering *S*'s susceptibility to trance and his ready motivation to respond to the hypnotist's demands. However, the sporadic nature of his imagery argues against the demand characteristic hypothesis. He failed to project some images externally and to evoke afterimages to imagined stimuli. Previous investigations have found that *S*s who are influenced by the demand characteristics of the experimental situation will report afterimages to imagined stimuli (7).

In conclusion, results indicate that *S*'s color responses to the TAT were a product of a genuine propensity for imaging which is quite atypical among adults in Western society. They highlight the special role cognitive capacities may play in determining subjects' responses to projective test stimuli.

E. Summary

During the course of an experiment, an *S* gave color responses to the TAT. The hypothesis was tested in the present study that *S* was an Eidetiker. Results indicated that he was able to evoke vivid eidetic images which persisted through time. Color responses to an achromatic stimulus could be explained by *S*'s unusual propensity for imagery, without any appeal to theorizing about the psychopathology of his adjustment. The results draw attention to the need to recognize the possible influence of cognitive factors on subjects' projective test responses.

References

1. Ammons, R. B., & Ammons, C. H. Quick Test. Missoula, Mont.: Psychol. Test Specialists, 1962.
2. Betts, G. H. The distribution and functions of mental imagery. *Contrib. Educ.* (*Teachers College*), 1909, No. 26.

3. Doob, L. W. Exploring eidetic imagery among the Kamba of central Kenya. *J. Soc. Psychol.*, 1965, **67**, 3-22.
4. ———. Eidetic imagery: A cross-cultural will-o'-the-wisp? *J. of Psychol.*, 1966, **63**, 13-34.
5. Haber, R. N., & Haber, R. B. Eidetic imagery: I. Frequency. *Percept. Motor Skills*, 1964, **19**, 131-138.
6. Hardy, L. H., Rand, G., & Rittler, M. C. AO H-R-R Pseudochromatic Plates. New York: Amer. Optical Co., 1957.
7. Hibler, F. W. An experimental investigation of negative after-images of hallucinated colors in hypnosis. *J. Exper. Psychol.*, 1940, **27**, 45-57.
8. Klopfer, B., Ainsworth, M. D., Klopfer, W. G., & Holt, R. R. Developments in the Rorschach Technique (Vol. 1. Technique and Theory). New York: World Book, 1954.
9. Murray, H. A. Thematic Apperception Test. Cambridge, Mass.: Harvard Univ. Press, 1943.
10. Nichols, J. Eidetic imagery synthesis. *Amer. J. Psychother.*, 1962, **16**, 76-82.
11. Orne, M. T., & O'Connell, D. N. Diagnostic ratings of hypnotizability. *Internat. J. Clin. & Exper. Hypnosis*, 1967, **15**, 125-133.
12. Rorschach, H. Psychodiagnostics. Berne: Hans Huber, 1942.
13. Rotter, J. B., & Rafferty, J. E. Manual: The Rotter Incomplete Sentences Blank, College Form. New York: Psychol. Corp., 1950.
14. Sheehan, P. W. Functional similarity of imaging to perceiving: Individual differences in vividness of imagery. *Percept. & Motor Skills*, 1966, **23**, 1011-1033 (Monog. Suppl. 6-V23).
15. ———. A shortened form of Betts' Questionnaire Upon Mental Imagery. *J. Clin. Psychol.*, 1967, **23**, 386-389.
16. ———. Artificial induction of posthypnotic conflict. *J. Abn. Psychol.*, in press.
17. Siipola, E. M., & Hayden, S. D. Exploring eidetic imagery among the retarded. *Percept. & Motor Skills*, 1965, **21**, 275-286.
18. Weitzenhoffer, A. M., & Hilgard, E. R. Stanford Hypnotic Susceptibility Scale, Form C. Palo Alto, Calif.: Consult. Psychol. Press, 1962.
19. Witkin, H. A. Individual differences in ease of perception of embedded figures. *J. Personal.*, 1950, **19**, 1-15.

Department of Psychology
The City College of The City
University of New York
Convent Avenue at 138th Street
New York, New York 10031

Published as a separate and in *The Journal of Psychology*, 1968, 68, 211-214.

INDIVIDUAL DIFFERENCES IN FIGURAL AFTEREFFECTS: SOME PROBLEMS AND POTENTIALS*

The University of Colorado and the Denver Veterans Administration Hospital

MICHAEL WERTHEIMER AND FRED H. HERRING

In 1951, Klein and Krech published a paper reporting that brain-injured people are more susceptible to kinesthetic figural aftereffects than are controls, and, among others, Eysenck (1) and Wertheimer (3) soon thereafter reported other personality, psychiatric, cognitive, and physiological correlates of figural aftereffects. In the decade between the early 1950's and the early 1960's, a number of laboratories concentrated on further study of individual differences in figural aftereffects, with the hope that such measured individual differences could reflect differences in some major underlying dimension, such as cortical conductivity (2), metabolic efficiency (3), or susceptibility to reactive inhibition (1). Two strategies appear to have guided these research efforts: first, the search for correlates of individual differences in figural aftereffects (other perceptual and cognitive functions, various personality and physiological indices, etc.), motivated largely by theoretical concerns; and, second, group comparisons (such as among psychiatric and neurological categories), with the aim of ultimately developing procedures which could be used for differential diagnosis.

During the late 1950's the senior author set up a figural aftereffects laboratory at Rockland State Hospital in New York, and during the early 1960's both authors developed a perception laboratory, which included a series of figural aftereffect devices, at the Denver Veterans Administration Hospital. Each of these laboratories[1] concentrated both on the theoretical correlational and the practical diagnostic problem, with a variety of perceptual functions (several visual and several kinesthetic figural aftereffects, ambiguous figure reversals, duration and fluctuation of negative afterimages, spiral aftereffects, critical flicker frequency, etc.) measured on a grand total of about 1000 per-

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 11, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The authors would like to express their gratitude to the following, who undertook many of the measurements or who facilitated the work of the laboratories in other ways: Glen Brennan, Arnold Greenhouse, Nathan Kline, James Lipton, John Means, Teresa Pressman, Robert Smilon, David Starrett, Nancy Wertheimer, and the Veterans Administration Western Research Support Center in Sepulveda, California.

sons in various categories (schizophrenics, patients with brain damage, neurotics, alcoholics, patients medicated with tranquilizers and energizers, controls, etc.) in about a dozen different studies.

Although a few reports were prepared (3, 4), most of the data the authors have gathered over the years have not been published. The present note is written in the hope that the experience of a decade and a half in this area may be of use to other investigators, and that others will thereby be more able to avoid some of the blind alleys and difficulties which have characterized much of the authors' work.

The typical finding, in correlational matrices of various figural aftereffect measures, other perceptual and cognitive indices, and personality data, has been that a larger number of correlation coefficients are in the expected direction, or are statistically significant, than would be anticipated by chance alone, but even the strongest correlations would be at a disappointingly low level (an *r* of the order of .30 would be atypically high). Frequently there were trends (sometimes not statistically significant), such as that schizophrenics exhibit lower susceptibility to the illusions, and brain-damaged patients greater (with one technique) or lower (with other techniques) susceptibility, or that medication with standard ataractic drugs reduced susceptibility. Isolated differences or correlation coefficients significant in one study would occasionally not be significant in a replication. In those cases in which significant correlation coefficients were obtained, it was sometimes possible that they might have been spuriously generated by group differences on the measures, due to contamination by the broad heterogeneity of the *S*s (e.g., schizophrenics, as one subgroup of *S*s in a larger *S* pool, may be less able than other *S*s to pay close attention to the procedures, resulting in lower scores on several otherwise unrelated perceptual measures).

As apparatus and procedures were developed so that several devices became available for the measurement of the "same" perceptual performance, it became evident that what at first seemed like relatively minute and unimportant differences in apparatus and procedure (such as the specific wording of the instructions, the physical setting within which the measures are taken, whether psychophysical judgments were made with or without counterbalancing of starting positions, the location of the POE in space relative to *S*'s body, whether an illusion was measured as an apparent contraction or an apparent expansion of a test object, the details of the design of the apparatus, the rapport between *E* and *S*) made major differences in the measured individual differences. Overall, measured individual differences in figural aftereffects appear to reflect a wide variety of as yet poorly understood, often subtle influences, many of which depend upon procedural differences among studies and investigators, which to date can only have served to obscure any possible underlying meaningful differences and correlations.

On the basis of the authors' experience, therefore, it would not appear to be very fruitful, at the present stage of technology of figural aftereffects, (*a*) to attempt extensive further explorations of individual differences in figural aftereffects of the kind extensively attempted so far, with the aim of developing relatively definitive diagnostic procedures; or (*b*) to engage in correlational work using present techniques with the hope of obtaining clear-cut data of major theoretical significance. The authors have come to this impression even though the findings to date have often been at least somewhat encouraging to such explorations; their conclusion is due to the fact that typical findings are at disappointingly low levels of significance or repeatability. The bewildering variety of apparatus and techniques for measuring individual differences in figural aftereffects has compounded the problem.

More appropriate, it seems, would be a concentration on fundamental psychophysical study of these perceptual functions. Perhaps it might eventually, then, become possible to develop more rational, standard procedures which could be designed to maximize any potentially meaningful individual differences. Enough nonchance findings exist already to suggest that an improvement in techniques, resulting in a reduction of the currently sizable "random" error, may eventually make it possible to obtain data of more clear-cut theoretical and diagnostic utility than has been achieved to date.

SUMMARY

On the basis of 15 years' experience with theoretical and diagnostic research on individual differences in figural aftereffects, it is suggested that it would be more fruitful to engage in fundamental dimensional psychophysical study of these perceptual phenomena than to continue further correlational efforts with the present variegated techniques. The order of significance, statistical, practical, and theoretical, of findings to date with the available methods is disappointingly low, and it appears likely that a standardized set of methods, once developed, may yield more satisfactory results.

REFERENCES

1. EYSENCK, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. Abn. & Soc. Psychol.*, 1955, **51**, 94-106.
2. KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain-injured. *J. Personal.*, 1952, **21**, 118-148.

3. WERTHEIMER, M. Figural aftereffect as a measure of metabolic efficiency. *J. Personal.*, 1955, **24**, 56-73.

4. WERTHEIMER, M., LIPTON, J. M., HERRING, F. H., GREENHOUSE, A. H., & MEANS, J. R. A re-examination of kinesthetic figural aftereffects in the brain-injured. *Percept. & Motor Skills*, 1965, **20**, 518-520.

Department of Psychology
The University of Colorado
Boulder, Colorado 80302

VA Hospital
1055 Clermont Street
Denver, Colorado 80220

Published as a separate and in *The Journal of Psychology,* 1968, **68,** 215-221.

THE LANGUAGE OF CHILDREN IN THE "INNER CITY"*[1]

Temple University[2]

THERON ALEXANDER, JUDITH STOYLE, AND CHARLES KIRK

A. PURPOSE

The complex task of acquiring language in the early years of life is, of course, a part of the response patterns significantly influenced by cultural requirements and conditions. Thus, information about language development has particular relevance for programs endeavoring to offset cultural and environmental adversity. Such programs must be concerned with language development if they are to ameliorate unfavorable influences and compensate for the lack of opportunities which are ordinarily part of the background of verbal experience essential for normal cognitive development.

Since language is learned from other people, deprived environments with familial dislocation and disorganization may not only fail to provide appropriate linguistic sex role models but may also fail to provide suitable reinforcement and varieties of environmental experience to bring about adequate verbal facility. Although general agreement exists that an adverse and deprived environment will impair children's development in language, there is as yet incomplete information about the nature of resulting inadequacies. Knowledge, for example, as to whether or not both boys and girls in a deprived environment develop similarly in rate and pattern and hence are able to benefit equally well from intervention, as is now being provided in the nationally supported Head Start programs, would be of value in curricular and experiential planning.

The very young children in this study come from some of the worst possible

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 11, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The research reported herein was performed pursuant to a contract with the Office of Economic Opportunity, Executive Office of the President, Washington, D.C. 20506. The opinions expressed herein are those of the authors, and should not be construed as representing the opinions or policy of any agency of the United States Government.

[2] Theron Alexander is the director of the Child Development Research Center for Head Start at Temple University in Philadelphia. The Center is one of the 13 national centers established to do research with programs for deprived children. This study was undertaken as part of the Center's research program supported by the National Office for Research and Evaluation for Head Start under the Office of Economic Opportunity and through the Institute for Educational Development. Judith Stoyle is a consulting statistician for the Center and Charles Kirk, now at the University of Maine, was formerly on the staff of the Center.

conditions of poverty in one of the world's largest cities. They play on garbage-strewn sidewalks beneath the broken windows of decaying "row houses" in a part of the city that has deteriorated steadily over the years. Not only do they live in conditions of "urban decay," but their family life is characterized by the many psychosocial factors of stress associated with frequent crises related to health, safety, and the unity of the family. The main purpose of this study is to seek information about the linguistic development of these children—information about sex differences in vocabulary and vocabulary change in reference to time and training.

The general question of whether or not sex differences in language development exist has been of interest to investigators for many years, although a consensus about a difference has not been attained. McCarthy (3, 4), in an evaluation of research, has asserted that sufficient evidence exists to conclude that girls are superior to boys in language development. She attributes the difference to child-training practices in which boys are encouraged to engage in active games and are sent out to play more often than girls. Accordingly, the superiority, in her view, grows out of the socialization process. Winitz (5), on the other hand, reports a study in which no sex difference in linguistic behavior was found in middle-class children in a midwestern university town. Still different findings were reported by Anastasi and D'Angelo (1) who found that Negro boys from nonracially mixed neighborhoods were superior to Negro girls on mean sentence length, while the white girls in the study surpassed the white boys. The authors were at a loss to explain their findings about the Negro boys in view of the general findings about the superiority of girls in most areas of language development. Although the issue of sex differences in language development is still unresolved after more than a quarter of a century of investigation, as methodologies are refined and as cultural variance is given much more significance, research problems at least now can be better defined.

The specific goals of the present study are (*a*) to determine if vocabulary differences exist between the sexes in children 3 to 5 years of age growing up in the deprived conditions of a large city; and (*b*) to determine if there is a significant vocabulary change over the Head Start school year and, if a change takes place, to determine if there is a difference between the sexes in the change.

B. METHOD

1. *Subjects*

The 118 children (52 girls and 66 boys) in this study are predominantly 3 and 4 years of age (see Table 1). No statistically significant differences in age exist between the boys and girls in the study. The subjects attended preschool (Head Start) in four centers located in socioeconomically depressed, urban areas having predominantly Negro residents.

TABLE 1
MEANS AND RANGES FOR AGES OF SUBJECTS AT BEGINNING OF STUDY

Sex	*N*	Mean	Range
Male	66	4.1	2.9—5.0
Female	52	4.2	3.2—4.8
Total	118	4.1	2.9—5.0

2. *Procedure*

The subjects in this study were given the Stanford-Binet Intelligence Scale as part of the national program for the study of children in preschool programs known as Head Start. The first testing of the children was done in the fall of 1966 and the second testing was done in the spring of 1967 with at least an interval of four months for all subjects. The data used in this study consisted of the scores of the subjects on the Picture Vocabulary items of the Stanford-Binet Intelligence Scale. In this part of the Binet, a subject is asked to identify 18 cards containing line drawings of objects. The pictures, drawn in black ink against a white background, show the following objects.

1. Airplane
2. Telephone
3. Man's hat
4. Ball
5. Tree
6. Key
7. Horse
8. Knife
9. Jacket
10. Ship
11. Umbrella
12. Foot
13. Flag
14. Cane
15. Arm
16. Pocket knife
17. Pitcher
18. Leaf

Each card is shown one at a time and the examiner asks: "What's this? What do you call it?" The picture vocabulary score is the sum of the correct responses given.

Since the normative data provided for the Binet are incomplete for each of the age levels covered in this study, normative scores where necessary were interpolated and are shown in Table 2. By using these normative levels, the test data can be a basis for comparing each child's performance against the normative criterion contained in the Binet (and interpolations).

TABLE 2
AGE NORMS FOR PICTURE VOCABULARY ITEMS BASED ON STANFORD-BINET AND INTERPOLATED VALUES

Age in months	Number of words expected at each age level
30—32	8
33—35	9
36—38	10
39—41	11
42—44	12
45—47	13
48—50	14
51—53	15
54—56	16
57—59	17
60—	18

C. RESULTS

The first goal of this study was to determine if vocabulary differences existed between the sexes. The results (see Table 3) indicate that significantly more males are equal to or are above the age norms than are girls. Thus, a significant sex difference in extent of vocabulary in favor of boys was encountered (chi square $= 3.37$, $df = 1$, $p < .05$).

TABLE 3
NUMBERS OF CHILDREN ABOVE AND BELOW PICTURE VOCABULARY NORMS[a]

Age norm comparisons	Number of male subjects	Number of female subjects	Total number of subjects
Equal to or above age norms	21	8	29
Below age norms	45	44	89
Total	66	52	118

[a] $\chi^2 = 3.37$; $p < .05$.

The second purpose was to determine whether or not a change in vocabulary occurred during the school year and whether or not a sex difference occurred in reference to this change. It was found that, while a significant change for the entire group was encountered ($t = 3.81$, one-tailed test, $p < .001$), no significant difference (see Figure 1) was found between the sexes in reference to changes (median test, chi square $= 1.0$, $df = 1$).

D. DISCUSSION

The finding of a sex difference in the responses to the Binet Picture Vocabulary is of interest in that, contrary to a number of studies in the past, boys exceeded girls in meeting tasks of the test. Most previous investigators of linguistic development have found either that girls excelled or that no sex difference existed. In this present study, however, approximately 32 per cent of the boys were equal to or exceeded the norm for their age, while only about 15 per cent of the girls were equal to or exceeded norms for their age. The basis for this difference, if substantiated by further studies of other samples, must be sought, as McCarthy has emphasized, during the period of infant development and in the very early experiences surrounding the child in his life space.

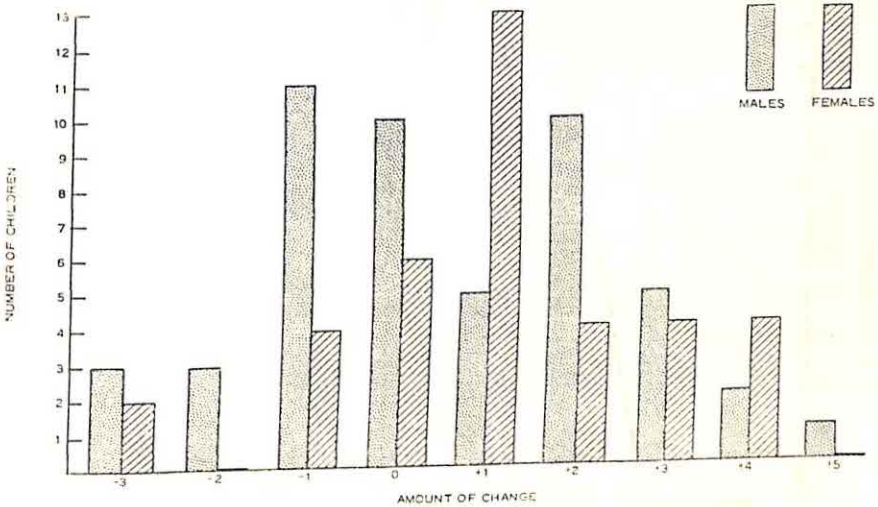


FIGURE 1

CHANGES IN INITIAL AND FINAL PICTURE VOCABULARY SCORES

Amount of change is in number of words passed. Mean change for entire group (mean of difference between first and second tests for each subject) = .759; $t = 3.81$; $p < .001$. Median difference in changes for boys and girls = 1.0. (There is no significant difference between the sexes in score changes over period of study.)

The existence of the sex difference should be considered, too, in relation to the fact that a large percentage of the group was below age norms (see Figure 2). The factors which have prevented these children from reaching age norms apparently have not interfered equally with both sexes and have been more devastating for girls than boys, particularly if, as has been asserted in the past, girls normally exceed boys in development. It should be noted that the finding of this study does not necessarily conflict with McCarthy's view of the accelerated development of girls as her studies apparently were based on the normal course of development and on middle-class children.

The present study supports the findings of Anastasi and D'Angelo that Negro boys in lower socioeconomic class exceed girls in the acquisition of linguistic skills. Hess and Shipman (2) have suggested that Negro mothers of lower socioeconomic class use less verbalization with their children than do middle socioeconomic class mothers. If girls are kept more with the mother or female substitutes, and the boys on the other hand are allowed to play more with other children, perhaps a difference occurs in this way.

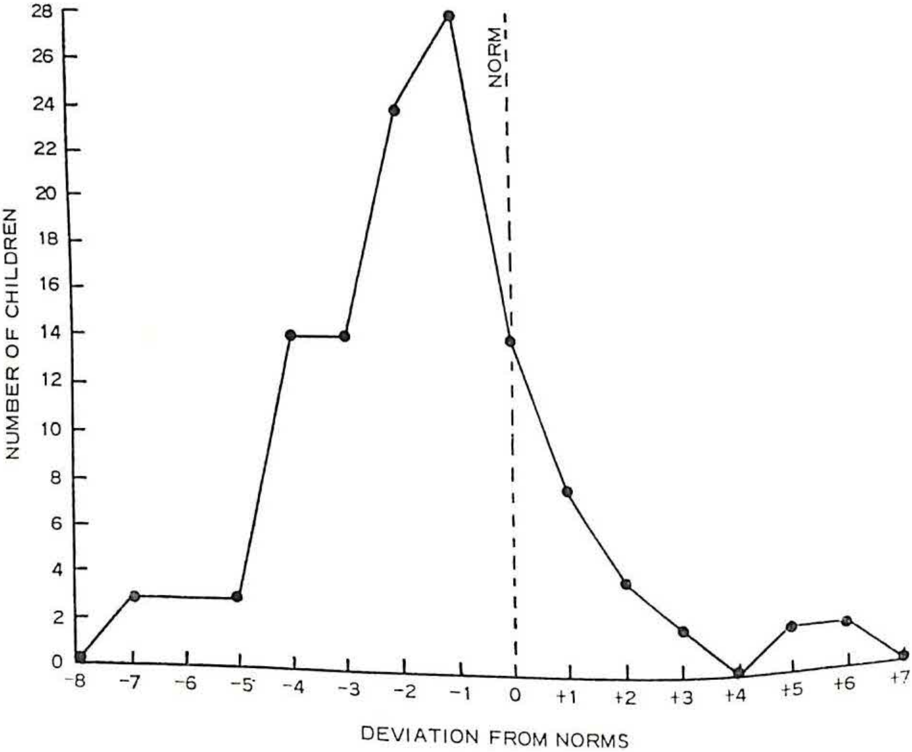


FIGURE 2
Picture Vocabulary Scores Compared to Stanford-Binet Norms for All Subjects

Of still other importance is the fact that the program of Head Start, with its purpose being to offset unfavorable intervening factors in development, did not alter the relative disparity between the sexes. Even though overall gain was achieved, the relative difference between the performance of the boys and the girls remained essentially the same. It would seem that, if the deficiency of the girls is to be altered, specific measures will be required.

E. SUMMARY

The subjects in this study (52 girls, 66 boys) are predominantly 3- and 4-year-old children living in dire conditions of urban decay. The investigation concerns picture vocabulary scores obtained at the beginning and end of the school year (Head Start). It was found that there was a significant sex difference in vocabulary in favor of boys. It was also found that, while a significant change in vocabulary was encountered over the school year for the entire group, the relative difference between the sexes was maintained. The information obtained suggests that, if the deficiency of girls in vocabulary is to be overcome, measures for specific help must be undertaken.

REFERENCES

1. ANASTASI, A., & D'ANGELO, R. Y. A comparison of Negro and white preschool children in language development and Goodenough Draw-A-Man I. Q. *J. Genet. Psychol.*, 1952, **81**, 147-165.
2. HESS, R. D., & SHIPMAN, V. C. Early experience and the socialization of cognitive modes in children. *Child Devel.*, 1965, **36**, 869-886.
3. MCCARTHY, D. Some possible explanations of sex differences in language development and disorders. *J. of Psychol.*, 1953, **35**, 155-160.
4. ———. Language development in children. In L. Carmichael (Ed.), *Manual of Child Psychology* (2nd ed.). New York: Wiley & Sons, 1954. Pp. 492-630.
5. WINITZ, H. Language skills of male and female kindergarten children. *J. Speech & Hearing Res.*, 1959, **2**, 377-386.

Child Development Research and
Evaluation Center for Head Start
Temple University
Philadelphia, Pennsylvania 19122

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 223-226.

TWO SCALES TO MEASURE COGNITIVE RIGIDITY*

The International Division of the American Institutes for Research;
and the Department of Educational Psychology,
The University of Calgary

JAMES G. SNIDER AND GRANT C. DRAKEFORD

A. INTRODUCTION

The purpose of this paper is to present two scales[1] for the measurement of cognitive rigidity and to list the results of a series of studies in which they have been used. Although the scales were originally intended to measure a variable called all-inclusive conceptualization, which was derived from various suggestions of general semantics (3, 4), recent factor-analytic data suggest that "cognitive rigidity" may be a better term for what these scales measure.

The scales are based on the arguments of a number of writers who have argued for the importance of a linguistic content variable which they termed "allness" and which was later called "all-inclusive conceptualization." The basic proposition was that absoluteness (allness) in language would be directly related to rigidity in behavior. Johnson (3), for example, suggested that a tendency toward all-inclusive conceptualizing may lead to various personal maladjustments. Korzybski (4) argued that all-inclusiveness is one of the major pitfalls of language which may lead to very general misevaluations of reality. Hayakawa (2) has made the observation that the individual who is too all-inclusive is likely to misjudge the environment, since he will not be flexible enough to change with a changing environment. The importance of the all-inclusiveness variable in personal counseling has been noted by Snider (8).

B. DEVELOPMENT OF THE SCALES

Cognitive rigidity was defined operationally as the tendency to respond in terms of absolute or overgeneralized language. It was taken as the tendency to prefer to use such terms as *all, always, never, and forever*. Although some researchers (5) have used word-counts of the use of these absolutistic terms, it seems preferable to develop scales to measure the tendency to use such

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 18, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] Copies of the scales are available from the authors.

terms, since through the use of scales we can control better for reliability and for standardization of stimuli.

The first cognitive rigidity scale (Word Preference Test) was based on the hypothesis that a tendency toward the use of absolutistic terms would be revealed by expressed preferences for one word or the other of pairs of words, where one of each pair is an absolutistic word and the other is not. It is assumed that in such a situation the person who tends to use absolutistic terms will tend to choose the absolutistic term in each word pair. Such pairs as "never-seldom" and "always-sometimes" were used.

Two versions of this type of scale have been constructed. Words are equated within pairs for frequency of use (13) to ensure that choices are not based on frequency of exposure to the stimuli. Each scale is composed of half "crucial" items—i.e., items containing absolutistic terms—and half "distractor" items, which are word pairs similar in meaning but which do not contain absolutistic terms. These items were introduced to ensure that subjects would not discover the purpose of the scales by simply scanning the items. Examples of "distractor" items are "den-cave" and "individual-person."

One version of this scale contains 80 items, while the other contains 60 items. Biserial item analysis was used in constructing the scales. Reliability figures (1st *vs.* 2nd half) were established on samples totaling more than 600 high school pupils and college students. These reliability figures range from .81 to .94. Validity studies were carried out by correlating Word Preference Test scores with word-counts of absolutistic terms used in free writing of high school pupils. Four validity coefficients range from .38 to .45 in samples totaling more than 200 pupils.

The second scale of cognitive rigidity (True-False Test) has two forms, each one containing 100 statements of the type "Teachers are strict," or "Ministers are good men," to which the subject is to respond true or false. The cognitive rigidity score is simply the total number of statements answered true. It is assumed that when agreeing with such statements the subject is exhibiting his degree of cognitive rigidity.

Items for the tests were carefully chosen from a large pool of items, and both forms of the test produced reliability coefficients (1st *vs.* 2nd half) of .90 to .94, calculated from several samples totaling more than 500 high school pupils and college students.

C. SOME CORRELATES OF THE COGNITIVE RIGIDITY SCALES

Several studies have been carried out using the various forms of the cognitive rigidity scales. The 80-item form of the Word Preference Test has been shown to correlate negatively with the flexibility scale of the California Psychological Inventory and positively with the conformance scale of that same inventory (6). It was also shown to correlate positively with a stereotyping test (7). The 60-item form of the Word Preference Test has been shown to be positively correlated with the Nufferno Level Test of nonverbal intelligence and the Rudin Response Set Scale (1).

A French translation of the 60-item form was shown to be positively correlated with high and low French milieus (12), suggesting some linguistic-cultural implications of cognitive rigidity. In one study, in which subjects were required to perform on a word preference task derived from the Word Preference Test, cognitive rigidity scores were found to be negatively correlated with intensity of discrimination between stimuli of high and low degrees of meaningfulness (10).

The True-False Test of cognitive rigidity has been found to discriminate between achievers and underachievers (11) and between delinquents and nondelinquents (9). It also has been shown to be negatively correlated with Otis Gamma *IQ* scores (9), and there is evidence that it is positively correlated with age but shows no difference by sex.

Recently Drakeford (1) investigated the relationship between various "overinclusiveness" tests and the cognitive rigidity measures by means of a factor analytic design consisting of a Principal Components analysis rotated to the Varimax criterion and computed for a matrix of 20 measures, composed of four measures of overinclusiveness, one form each of the cognitive rigidity scales, and 14 "marker" variables.

The data showed no relationship between overinclusiveness and cognitive rigidity, but the measures of cognitive rigidity did show cohesion in that they both loaded on a common factor which Drakeford labeled "cognitive rigidity" and which was loaded by the Rudin Response Set Scale, the CPI Flexibility Scale, the Rokeach Dogmatism Scale, the "R" Scale of Acquiescence, and the two cognitive rigidity scales.

The latter study suggests that the cognitive rigidity measures may be measuring a general unidimensional variable which is loaded in the direction of "rigidity" and which is cognitive in nature.

D. Summary

Two scales are presented which derive from the linguistic content variable of "allness" but which are shown in a factor analysis to load on "cognitive rigidity." Results of several studies are listed which show the scales to be useful measures of cognitive rigidity. Copies of the scales are available from

the authors. All forms of the scales, except one form of the True-False Test, may be seen in Snider (6) or Drakeford (1).

REFERENCES

1. DRAKEFORD, G. C. Cognitive and linguistic inclusiveness: A factor analysis. Unpublished Master's thesis, University of Calgary, Calgary, Canada, 1967.
2. HAYAKAWA, S. I. Language in Thought and Action. New York: Harcourt, Brace, 1943.
3. JOHNSON, W. People in Quandaries. New York: Harpers, 1946.
4. KORZYBSKI, A. Science and Sanity. Lakeville, Connecticut: Inst. Gen. Semantics, 1948.
5. OSGOOD, C. E., & WALKER, E. E. Motivation and language behavior: A content analysis of suicide notes. *J. Abn. & Soc. Psychol.*, 1959, **59**, 58-67.
6. SNIDER, J. G. Some correlates of all-inclusive conceptualization in high school pupils. Unpublished Doctoral dissertation, Stanford University, Stanford, California, 1964.
7. ———. Stereotyping and the use of all-inclusive terms. *Albert. J. Educ. Res.*, 1964, **10**, 170-174.
8. ———. The linguistic content variable in school counseling. *Person. Guid. J.*, 1964, **42**, 577-580.
9. ———. A comparison of all-inclusive conceptualization in the delinquent and non-delinquent. *Can. J. Corrections*, 1965, **7**, 411-413.
10. ———. All-inclusive conceptualization and intensity meaningfulness: A research note. *Albert. J. Educ. Res.*, 1966, **4**, 281-284.
11. SNIDER, J. G., & DRAKEFORD, G. C. All-inclusive conceptualization in academic achievers and underachievers. *Psychol. Schools*, 1967, **4**, 172-173.
12. SNIDER, J. G., & LAMBERT, R. A. Relative intensity of expression in Eastern and Western French-speaking Canadians. Unpublished manuscript, University of Calgary, Calgary, Canada, 1967.
13. THORNDIKE, E. L., & LORGE, I. The Teacher's Word Book of 30,000 Words. New York: Columbia University, 1944.

The International Division of the
American Institutes for Research
135 North Bellefield Avenue
Pittsburgh, Pennsylvania 15213

Department of Educational Psychology
The University of Calgary
Calgary, Alberta, Canada

Published as a separate and in *The Journal of Psychology,* 1968, 68, 227-242.

AN ANALYSIS OF MULTITRIAL FREE RECALL LEARNING WITH INPUT-ORDER HELD CONSTANT*

Department of Psychology, Vassar College

WILLIAM A. ROBERTS

A. INTRODUCTION

In several recent papers, it has been suggested that the typical bow-shaped serial position curve (SPC) of free recall learning (FRL) could be understood in terms of a two-process theory of memory (3, 10). Recall of late input-items is attributed to output from short-term storage, while recall of early and middle input-items is attributed to output from long-term storage. In this paper, a two-process model of FRL is elaborated, and it is shown that this model can account for a good deal of data in FRL. An experiment is presented which tests some predictions from this model.

The SPC of FRL typically demonstrates a marked recency effect which often is accompanied by early recall of terminal input-items. If this phenomenon represents an early unloading of items in short-term storage before they are lost, any factors which prevent the immediate recall of items in short-term storage should lead to the loss of these items. In support of this prediction, both Glanzer and Cunitz and Postman and Phillips have found that delaying recall for 30 seconds eliminated the recency effect but produced little change in the primacy effect (3, 4). Both Raffel and Deese required *S*s to recall items in the same order as they were presented (1, 5). This procedure, which effectively blocks the early recall of late input-items, led to a significant reduction in the recency effect as compared with that obtained under free recall instructions. The findings of an experiment by Tulving and Arbuckle are also clearly in line with this prediction (9). Tulving and Arbuckle presented *S*s with 10 digit-word pairs and then asked *S*s to respond with the correct word when a stimulus-digit was presented. Stimulus-digits at each input-position were presented systematically at each output-position over trials so that recall could be plotted as a function of output-position with input-position as a parameter. The curve for the final two input-positions shows a high degree of recall at the first two output-positions followed by a marked

---

* Received in the Editorial Office, Provincetown, Massachusetts, on December 26, 1967, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

decline at the later output-positions. Curves for the early and middle input-positions are flat, being invariant with output-position.

The model holds that items occurring in early and intermediate input-positions are recalled from a long-term storage system which is affected little by delays in recall. However, the primacy effect requires such a theory to explain why more early input-items are recalled from long-term storage than intermediate input-items. The primacy effect can be dealt with by an extension of the *differential-rehearsal-hypothesis,* early suggested by Welsh and Burnett (11). This hypothesis holds that early input-items are rehearsed more than later input-items, and, therefore, are remembered better. Why are early items more susceptible to rehearsal than later items? It is hypothesized that early input-items enter a short-term storage system where rehearsal of items allows the *S* to *organize* or *unitize* these items into long-term storage units. Short-term storage could be thought of as preorganized memory in which each item is remembered as an individual unit. Items pass into long-term memory when they are organized into superordinate structures. During input, *S* holds items just presented in short-term storage while he rehearses items presented earlier. The longer an item is kept in short-term storage, the greater is the possibility it will be lost from memory and not be rehearsed at all. Early input-items then have the advantage because they do not have to be held in short-term storage before being rehearsed. As the middle of the list is approached, however, more and more items must be held in short-term storage while earlier ones are rehearsed, and many of these intermediate items are lost from short-term storage. Thus, the earlier an item is presented, the greater is the probability that it will be rehearsed and pass into long-term storage; and, the later it is presented, the greater is the probability that it will be lost before there is opportunity to rehearse it.

There is evidence that relative usage of these two storage systems varies with the conditions of learning. Conditions which favor organization lead to relatively more and earlier recall from long-term storage than from short-term storage, and conditions which do not favor organization lead to relatively more and earlier recall from short-term storage than from long-term storage. Thus, if long-term memory is highly organized, output from long-term storage will be primary in recall; but, if few items enter long-term storage because the material is very dificult to organize, output from short-term storage will be primary in recall.

Variables which we would expect to affect degree of list-organization produce clear changes in the SPC. Free recall of low meaningful material, nonsense syllables, yields a SPC with a marked recency effect and little primacy effect (5, 11). Sumby directly compared the free recall of low and high frequency word-lists (6). The first trial SPC for the high frequency list shows equivalent primacy and recency effects. The SPC for the low frequency list demonstrates a recency effect equivalent to that for the high frequency list, but no primacy effect whatsoever. Increasing meaningfulness raises the level of the early portion of the SPC but has little effect on the terminal portion of the curve.

Deese and Kaufman studied the FRL of lists of words varying in linguistic structure from random ordering of words (zero order) to text (2). SPCs for each order of approximation to normal structure showed a clear trend from a recency effect and no primacy effect at zero order to a primacy effect and no recency effect at text. This finding follows from the model, since the nearer list-structure comes to normal language structure, the more items *S* should be able to organize into a single unit for long-term storage; with text, *S*s may unitize entire phrases or sentences. The fact that no recency effect was found with text suggests that output from long-term storage may have dominated recall completely. This suggestion is supported by the finding that *S*s recalled the text-list in the order it was presented to them. For a zero order list, order of recall was quite different, terminal items being recalled quite early with initial items being recalled next, somewhat sooner than intermediate items. When opportunity for organization is relatively low, output from short-term storage appears to be primary in recall. Some of Sumby's results also are in agreement with this finding (6). When *S*s were required to recall a list of high frequency words, they emitted early input-items first. When *S*s recalled a low frequency list, however, output-order was reversed with late input-items coming first.

In two early experiments, Welsh and Burnett and Raffel instructed *S*s to memorize each nonsense syllable of a list as it was presented but not to think of an item after it had been presented (5, 11). A set not to rehearse which prevents *S*s from relating items to one another should strongly inhibit organization. The SPC for this condition yielded a marked recency effect but absolutely no primacy effect.

Finally, the model predicts that providing *S*s with opportunity for rehearsal should facilitate the primacy effect more than the recency effect. Glanzer and Cunitz found that spacing input-items in time significantly aided the recall of early and intermediate input-items but did not affect the recall of terminal items (3). Repeated trials should also provide opportunity for organization. Sumby found that, after three trials, the SPC for a low frequency list showed equivalent primacy and recency effects (6). Since the first trial SPC showed a recency effect but no primacy effect, learning between Trials 1 and 3 must have proceeded at a faster rate for early items than for late items.

In the experiment reported here, degree of list-organization is varied by requiring *S*s to recall lists of categorized and uncategorized words. Each list is learned for 10 trials with the input-position of words remaining constant over trials. The two-process model makes several predictions regarding this experiment. The Categorized List should show superior recall to the Uncategorized List, since the former provides common conceptual categories for organization while the latter does not. Further, the benefits of categorization should be more marked at the beginnings and middles of the lists than at the ends, since it is held that organization facilitates the passage of early and intermediate items from short-term to long-term storage; final items should be about equally well recalled from short-term storage in both lists.

Tulving has shown that subjective organization of a list of words increases over trials in FRL even though input-order is varied from trial to trial (8). An experiment where input-order is held constant over trials allows one to examine subjective organization at different points within a serial list. The differential-rehearsal-hypothesis holds that opportunity for organization through rehearsal diminishes from beginning to end of a list, and, thus, it is predicted that subjective organization should decrease as a function of serial position.

As successive trials provide repeated opportunities for rehearsal, long-term storage should become enlarged and strengthened relative to short-term storage. The two-process model then suggests that early items might be learned at a faster rate than terminal items, as Sumby found for low frequency words. It also might be expected that there would be a shift in primacy of recall over trials toward relatively earlier recall of initial input-items. This follows from Sumby's (6) and Deese and Kaufman's (2) finding that the more easily a list can be organized, the earlier initial input-items are recalled.

B. METHOD

The Categorized and Uncategorized lists each contained 45 words drawn from the Thorndike and Lorge list; the frequency for all words was five or less per million in the general count (7). The Categorized List was made up of 15 names of animals, 15 articles of clothing, and 15 men's and women's first names. Fifteen triads were formed, each containing one word from each category, with the order of items within triads randomized. Fifteen lists were constructed with each triad occurring once at each input-block of three words. Uncategorized words were chosen to be as unrelated to one another as possible. These words were randomly divided into triads, and 15 lists were constructed, again balancing triads over input-blocks.

The *S*s were 15 undergraduate women enrolled at Vassar College. One of each of the 15 categorized and uncategorized lists was learned by each *S*, eight *S*s learning the Categorized List first and seven *S*s learning the Uncategorized List first.

When *S* entered the laboratory, she was told that a series of words would be presented to her and that she was to memorize as many of them as possible; following presentation, she was to write down as many words as she could remember in any order she wished to recall them. The list was presented on a memory drum at the rate of one word per second. A number preceded the first word and followed the last word, signaling the beginning and end of the list. As soon as *S* saw the final number, she began to write on a lined recall sheet provided by *E*. A four-minute period was allowed for recall, the next trial beginning immediately after the end of this period. Following the tenth trial on the first list, *S* was excused from the laboratory and requested to return to the laboratory to learn the second list on another day.

C. RESULTS

1. *Words Recalled*

In Figure 1, it can be seen that the learning curves for both the Categorized and Uncategorized lists take the negatively accelerated form typical of FRL. The difference between the lists on Trial 1 appears to remain fairly constant over trials, the curves rising in a parallel fashion. A Lists × Trials × *S*s analysis of variance shows that the difference between lists is significant ($F = 23.13$, $df = 1/14$, $p < .001$), the increase in words recalled over trials is significant ($F = 317.37$, $df = 9/126$, $p < .001$), but Lists and Trials do not significantly interact ($F = 1.21$, $df = 9/126$, $p > .05$). Categorized words are recalled better than uncategorized words, and the advantage of categorization remains constant over the 10 trials studied.

The two-process model predicts that the difference between lists should be greater at the beginnings and middles of the lists than at the ends. In Figure 2, intratrial curves are presented for Trial 1 and successive groups of three trials. These curves plot words recalled as a function of input-blocks of nine words. It can be seen that, in all four sets of curves, the biggest differences between lists occur at the first four input-blocks, the curves tending to converge at the fifth block. The differences between lists in total words recalled over all 10 trials were tested at each input-block. The differences for the first four input-blocks are all significant (t's $= 1.83$, 2.18, 2.70, and 3.78 for Blocks 1-4, respectively, $df = 14$, $p < .05$ in all cases for a one-tailed test). The difference at the fifth block, however, is not significant ($t = 1.51$, $df = 14$, $p > .05$). The predicted relationship between lists and input-position is supported.

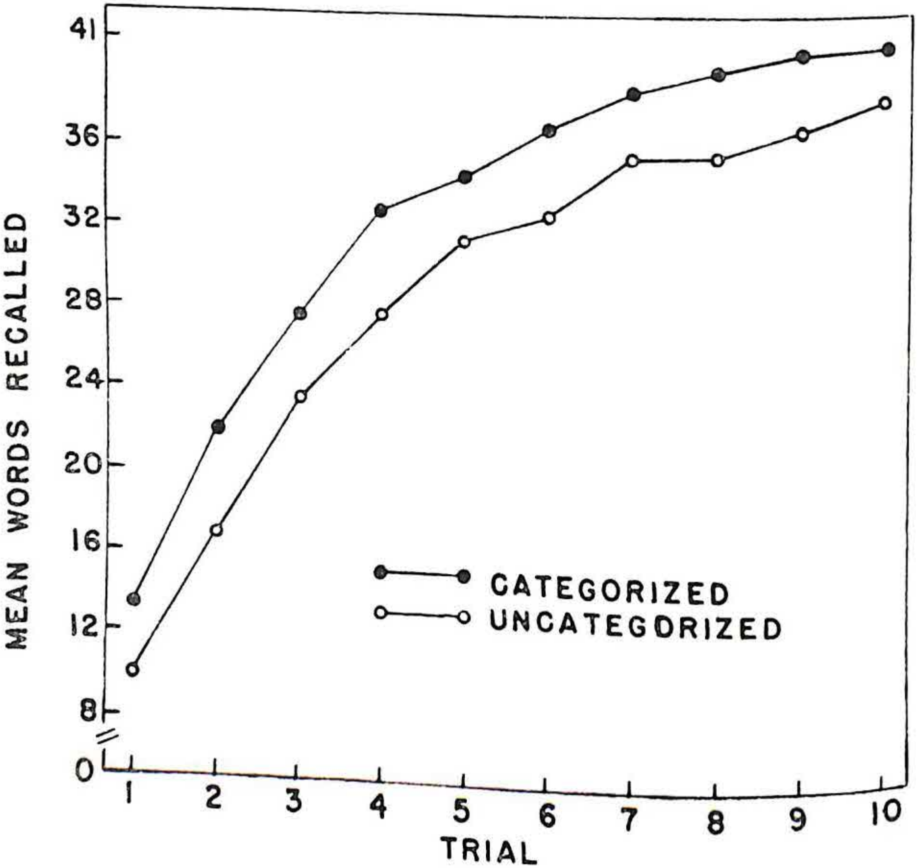


FIGURE 1
LEARNING CURVES FOR THE CATEGORIZED AND UNCATEGORIZED LISTS

Figure 2 shows that the least words were recalled at the fourth input-block in both lists, usually followed by a sharp rise in recall at the fifth block. By Trials 8-10, this effect has disappeared in the Categorized List, but there is still a clear drop in recall at the fourth block in the Uncategorized List. If the storage-limit of short-term memory is about eight to 10 units, as has been suggested, then words in the fourth input-block would not be recalled from short-term storage. Coming late in the list, they also have little opportunity to be rehearsed and enter long-term storage. The fourth block, then, should be the most disadvantageous position in the list according to the two-process model. The fact that the difficulty of recall at Block 4 eventually is overcome in the Categorized List but not in the Uncategorized List is also in line with the model, since categorized words at Block 4 should require less rehearsal than uncategorized words in order to be organized into long-term storage units.

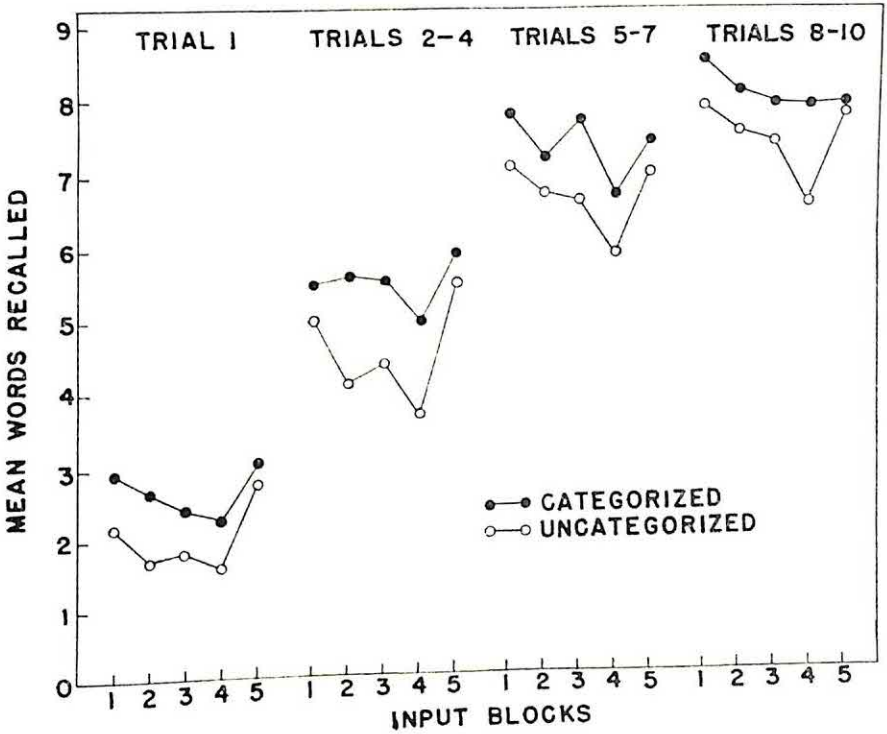


FIGURE 2
WORDS RECALLED AT EACH INPUT-BLOCK AT SUCCESSIVE STAGES OF LEARNING FOR THE CATEGORIZED AND UNCATEGORIZED LISTS

Finally, there is a shift over trials in the relative degree of recall from the beginnings and ends of the lists. Recall is somewhat better at the beginnings of the lists than at the ends in the later trials, whereas the opposite is the case in the earlier trials. While this effect is not marked, it will be seen that it is in agreement with the order of recall functions.

2. *Order of Recall*

For each *S*, the mean ordinal output-position of words recalled from each input-block was calculated at each trial. The means of these values are plotted at different stages of learning in Figure 3. In the early trials, terminal input-items are recalled sooner than early and intermediate input-items. As trials progress, words from the first input-block come to be recalled more and more early relative to words from other input-blocks. In the late trials, words from the final input-block continue to be recalled earlier than those in the intermediate blocks, but their initial primacy over Block 1 words has been lost; in fact, the primacy of terminal items over intermediate items in the Categorized List has been almost completely eliminated by Trials 8-10.

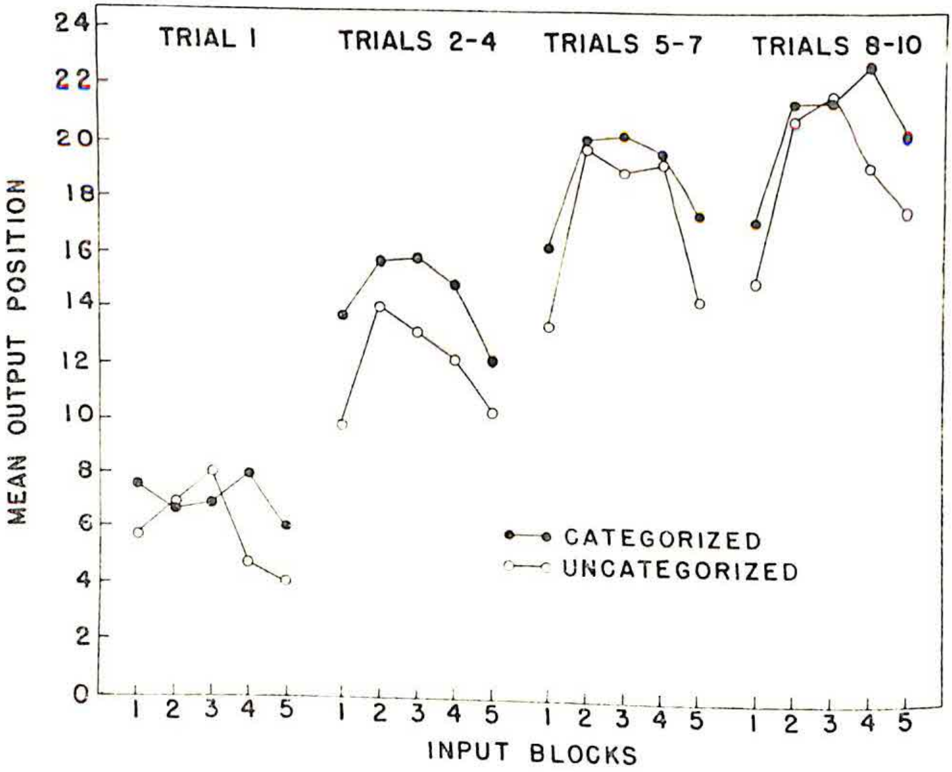


FIGURE 3
MEAN ORDINAL POSITION IN RECALL PLOTTED AS A FUNCTION OF INPUT-BLOCKS AT SUCCESSIVE STAGES OF LEARNING FOR THE CATEGORIZED AND UNCATEGORIZED LISTS

3. *Measures of Organization*

Two measures of organization were examined. The first, ratio of repetition (RR), was used to determine the degree of category-clustering occurring on successive trials in recall of the Categorized List. Mean RR is plotted against trials in Figure 4, and it can be seen that RR rises sharply from initial values of about .50 to an asymptotic level of about .69. Although input-position remained constant over trials, there is clearly a progressive organization of words into category-clusters.

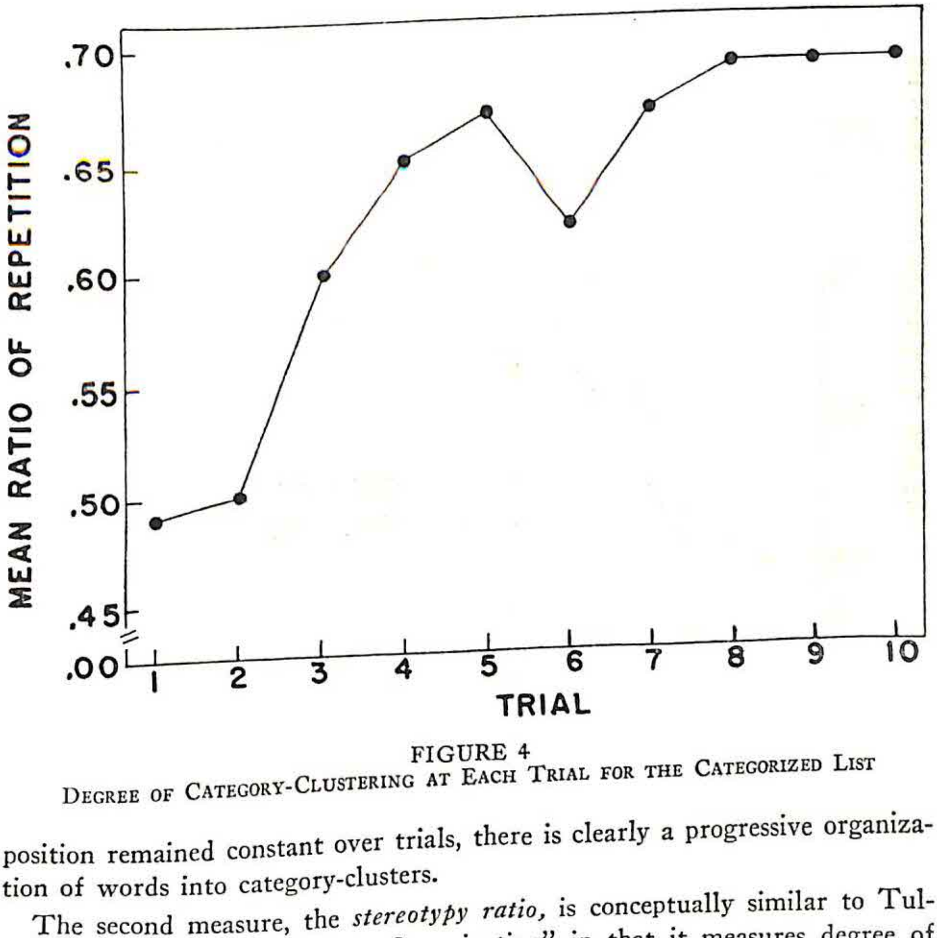


FIGURE 4
DEGREE OF CATEGORY-CLUSTERING AT EACH TRIAL FOR THE CATEGORIZED LIST

The second measure, the *stereotypy ratio,* is conceptually similar to Tulving's measure of "Subjective Organization" in that it measures degree of intertrial stereotypy in recall order for successive pairs of trials (8). If words recalled in Positions n and n + 1 on Trial 1 are recalled in adjacent positions (either in the same or in reversed order) on Trial 2, one opportunity for repetition is placed in the denominator of the ratio and one repetition is placed in the numerator. If words in Positions n and n + 1 on Trial 1 appear on Trial 2 but are not adjacent, one opportunity for repetition is placed in the denominator, but nothing is entered in the numerator. Finally, if one or both of the words in Positions n and n + 1 on Trial 1 fail to appear on Trial 2, nothing is entered in either the denominator or numerator. For all of the successive pairs of words recalled on Trial 1 by a given *S*, the frequency of repetition divided by the frequency of opportunity for repetition yields the stereotypy ratio. This value is calculated for each of the nine successive pairs of trials. Figure 5 presents the mean stereotypy ratio plotted as a function of pairs of trials for both categorized and uncategorized lists. Both curves show a rise in stereotypy from initial ratios of about .25 to ratios around .40 on the last three pairs of trials. An analysis of variance

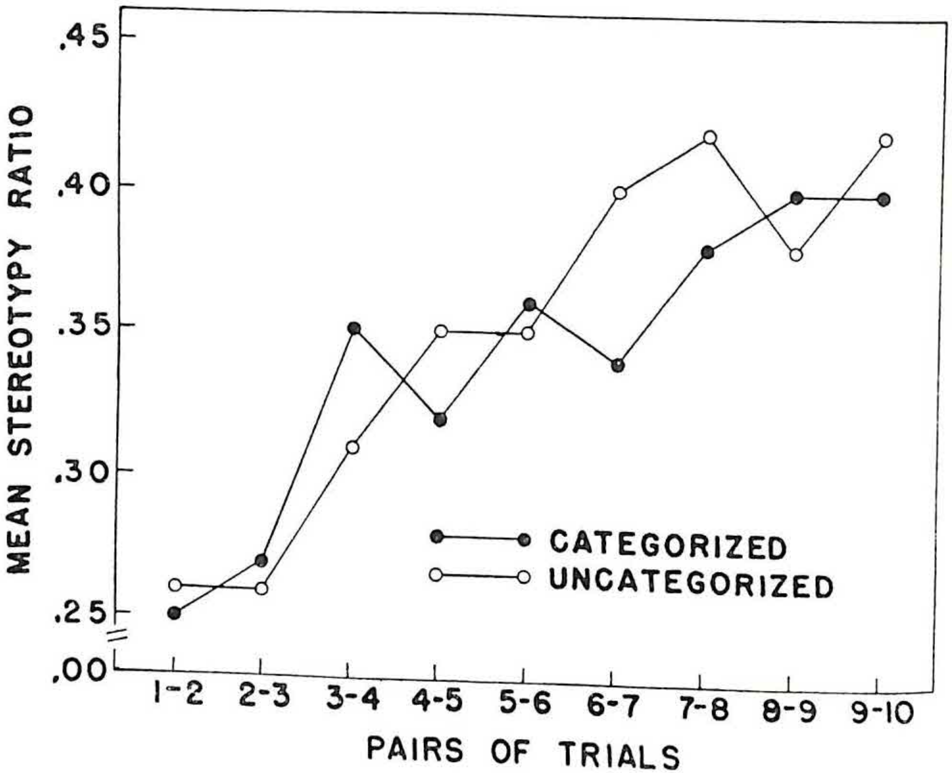


FIGURE 5
DEGREE OF INTERTRIAL STEREOTYPY IN ORDER OF RECALL AT SUCCESSIVE PAIRS OF TRIALS FOR THE CATEGORIZED AND UNCATEGORIZED LISTS

applied to these curves shows that the increase in stereotypy ratio over trials is significant ($F = 7.59$, $df = 8/112$, $p < .001$), but neither the difference between lists nor the interaction of Lists with Trials approaches significance ($F < 1.00$ in both cases). Thus, the imposition of rigidity upon order of recall seems to progress at an equal rate for categorized and uncategorized lists. Tulving has suggested that the recall of words in category-clusters may involve less rigid organization than is the case with "unrelated" words (8, p. 352). The data presented in Figures 4 and 5 suggest that, while there is increasing recall of categorized words in clusters over trials, the order in which words within clusters are recalled becomes increasingly fixed.

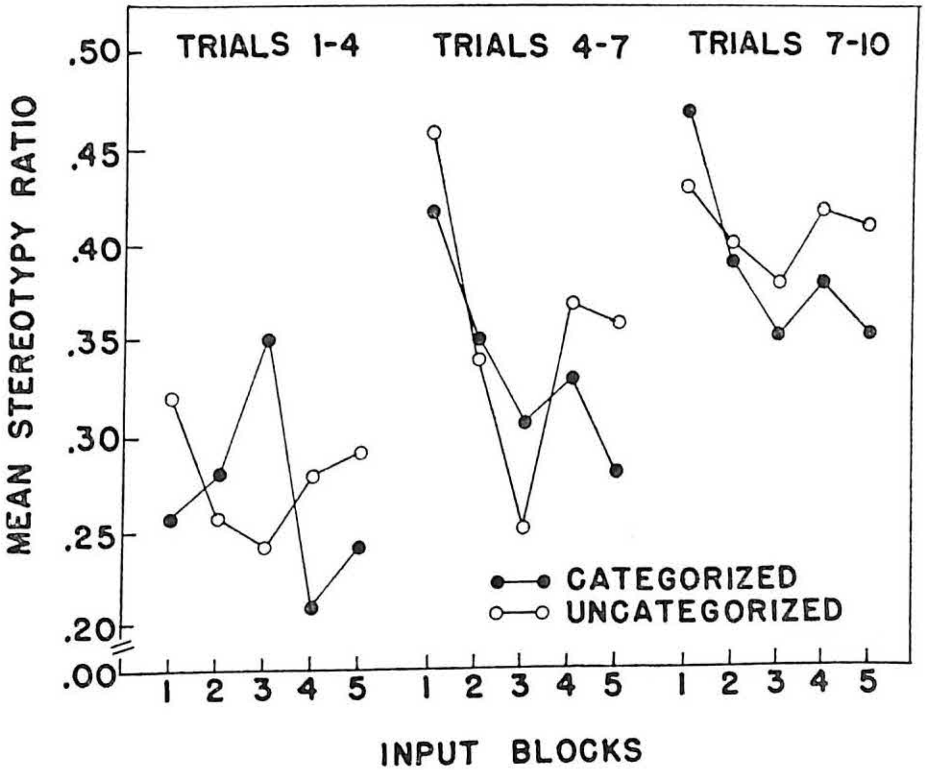


FIGURE 6
MEAN STEREOTYPY RATIO PLOTTED AS A FUNCTION OF INPUT-BLOCKS AT SUCCESSIVE STAGES OF LEARNING FOR THE CATEGORIZED AND UNCATEGORIZED LISTS

In order to study the degree of organization imposed on items recalled from different input-positions, separate stereotypy ratios were calculated for words recalled from each of the five input-blocks. Items from any given input-block were treated as items in Position n + 1, and the relative frequency with which they were repeated adjacent to whatever item was in Position n was calculated. These ratios were averaged over pairs of trials encompassing Trials 1-4, 4-7, and 7-10, and they are plotted for categorized and uncategorized lists in Figure 6. With the exception of the unusual rise in stereotypy at Block 3 on Trials 1-4, the curves for the Categorized List show the highest degree of organization at the initial input-positions, followed by a decline at the later positions. The curves for the Uncategorized List are clearly different. They also show the highest ratios at the first input-block and then decline. After reaching a minimum at Block 3, however, they show an abrupt rise at Blocks 4 and 5. While the curves for the Categorized List agree with the notion that differential-rehearsal should lead to a progressive decline in organization from the beginning of a list to the end, those for the Uncategorized List do not.

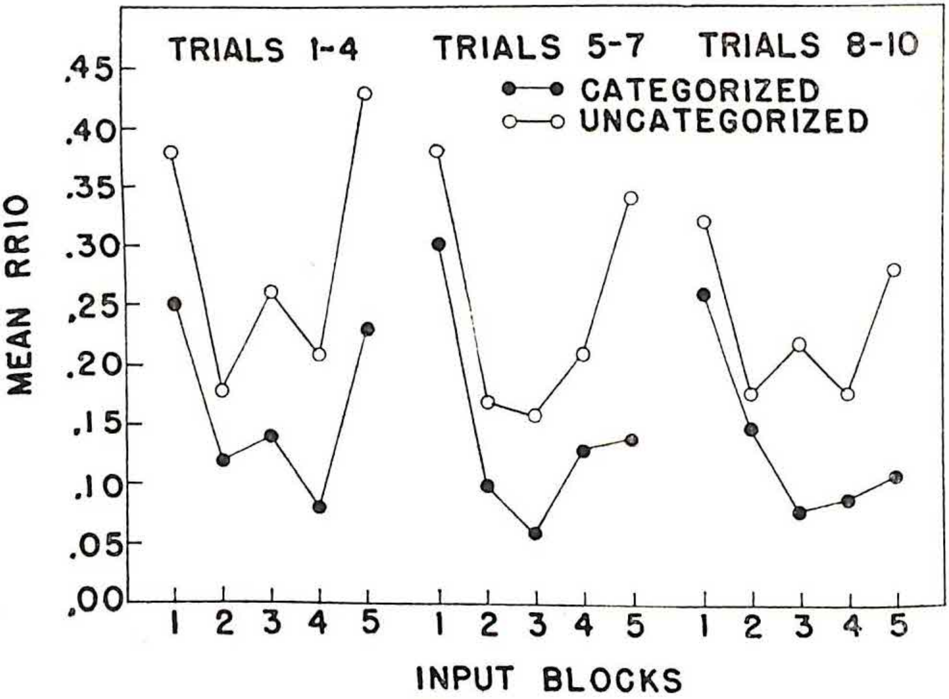


FIGURE 7
THE RATIO OF RECALL IN INPUT-ORDER PLOTTED AS A FUNCTION OF INPUT-BLOCKS AT SUCCESSIVE STAGES OF LEARNING FOR THE CATEGORIZED AND UNCATEGORIZED LISTS

One factor which could explain the rise in stereotypy ratio at the end of the Uncategorized List is a tendency for *S*s to recall terminal input-items from short-term storage in the order in which they are presented. Repeated recall in presentation-order would fix the order of recall and raise the stereotypy ratio.

4. *Recall in Presentation Order*

In order to test the above hypothesis, the *Ratio of Recall in Input Order* (RRIO) was calculated at each input-block. The opportunities for recall in presentation-order were determined by counting the number of pairs of items recalled which had occurred in Positions n and n + 1 during presentation. This frequency was divided into the frequency with which Items n and n + 1 were actually recalled in adjacent positions. Mean RRIO curves are plotted for Trials 1-4, 5-7, and 8-10 in Figure 7. Recall in presentation order is clearly more often the case in the Uncategorized List than in the Categorized List, and this is not unexpected, since category-clustering usually precludes recall in input-order. The curves for Trials 1-4 show a marked increase in the usage of input-order at the first and last input-blocks relative to the middle three blocks. The terminal rise in RRIO has dropped completely out of the Categorized Curve by Trials 8-10, while it remains, somewhat attenuated, in the Uncategorized Curve.

D. DISCUSSION

Results for the Categorized List agree well with the two-process model. Terminal categorized words have an advantage over words in the middle of the list on the early trials in that they are recalled better, recalled earlier, and tend to be recalled more often in order of presentation than intermediate words. This initial advantage of terminal items diminishes over trials until, in the final trials, terminal words are recalled no better than intermediate words, recalled no earlier than intermediate words, and are no more often recalled in presentation order than intermediate words. Initial input-items, on the other hand, show on the early trials an advantage in these measures which is maintained or increased in the later trials. This pattern of findings strongly suggests that terminal items are recalled from short-term storage on early trials, but, as trials proceed, many of them are pulled into organized clusters formed of earlier input-words. Initial input-words, having the advantage of extra rehearsal, early may be formed into strong long-term storage units which act as anchors for clusters. The stereotypy curves shown in Figure 6 further support the hypothesis that the Categorized List is organized from beginning to end, since these curves generally show a decline in stereotypy over serial positions.

A different pattern of results appears for the Uncategorized List. Throughout all stages of learning, there is a marked decline in words recalled at Block 4 followed by a sharp rise at Block 5. This clear superiority in recall of Block 5 over Block 4, along with the rise in primacy of recall, stereotypy, and recall in input-order at Block 5, suggests that *S*s were making extensive use of short-term storage information throughout learning of the Uncategorized List. It seems quite possible, in fact, that the repeated recall of short-term storage items in presentation-order may have served as a form of rehearsal which allowed many terminal items to enter long-term storage. If this is the case, the beginning and end of the Uncategorized List may have been organized independently into long-term storage units, the former through differential opportunity for rehearsal during presentation and the latter through continued recall of words from short-term storage in a stereotyped order. The differences between the Categorized and Uncategorized lists can then be understood in terms of different patterns of organization. In the Categorized List, terminal items are integrated with earlier items under a unitary scheme of organization. In the Uncategorized List, terminal items are organized independently of earlier items through their continued recall from short-term storage. The validity of this interpretation can be evaluated only by further experiments.

In general, the predictions made from the two-process model about multitrial FRL with fixed input-order were borne out. SPCs did converge at the final input-block, and there was a shift over trials toward greater and earlier recall of initial items relative to terminal items. Further, subjective organization, as measured by the stereotypy ratio, was higher at the beginning of a list than at the later positions, as the differential-rehearsal-hypothesis predicts. One aspect of these findings does not agree with this general theory. The poor recall and low RRIO at Block 4 of the Uncategorized List suggest that subjective organization should be low at this block and, yet, the stereotypy curves show a consistent rise from Block 3 to Block 4. In fact, the stereotypy ratio is as high at Block 4 as it is at Block 5. One possible explanation of this discrepancy is that *S*s, realizing that there is little opportunity to rehearse words in Block 4, adopt a strategy of selecting only a few words for rehearsal, while ignoring the rest. Such a strategy would lead to a low level of recall but a high level of stereotypy. Further investigation clearly is needed on this problem.

One other finding of interest in these data is the persistent tendency of *S*s to recall initial input-items in presentation-order. One indication of the strength of this tendency is the fact that RRIO remains high at Block 1 for the Categorized List, even on trials 8-10 where Figure 4 shows that the list in general has been highly organized into category-clusters. While this effect might be due to a stronger tendency to rehearse words in input-order at the beginning of the list than in the middle, it can be accounted for as a simple effect of differences in opportunities for reorganization. When few words have been presented, the number of new arrangements of words possible is relatively small. As more words are presented, the number of opportunities for new groupings of words expands rapidly, and the frequency of rehearsal of words in input-order should decline. Until evidence to the contrary is presented, this interpretation seems to be the most parsimonious account of this finding.

E. SUMMARY

Fifteen *S*s learned a list of 45 categorized words and a list of 45 uncategorized words for 10 trials by the free recall method with input-order held constant. Analysis of the data yields the following findings: (*a*) The Categorized List is recalled better than the Uncategorized List, but this difference is significant only at the early and intermediate serial positions, SPCs tending to converge at the terminal serial positions. (*b*) Words from the end of a list are recalled first on the initial trials, but, as trials proceed, there is a shift in primacy of recall with words at the beginning of a list being recalled first. (*c*) Measures of clustering and stereotypy of recall order show that both lists become progressively more organized over trials. (*d*) Stereotypy of recall order is greater for items in the initial serial positions than for items from later serial positions. (*e*) Words at the beginning and end of a list are more often recalled in the order in which they are presented than are words from the middle of a list. (*f*) Differing patterns of results for the Categorized and Uncategorized lists suggest that all items in the Categorized List are pulled into a unitary scheme of organization, whereas the beginning and end of the Uncategorized List are organized independently of one another.

These results, along with those of other FRL experiments, can be accounted for by a two-process theory of FRL which holds that items in initial and intermediate serial positions are recalled from a long-term storage system and that items from terminal serial positions are recalled from a short-term storage system.

References

1. Deese, J. Serial organization in the recall of disconnected items. *Psychol. Rep.*, 1957, **3**, 577-582.
2. Deese, J., & Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. Exper. Psychol.*, 1957, **54**, 180-187.
3. Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *J. Verb. Learn. & Verb. Behav.*, 1966, **5**, 351-360.
4. Postman, L., & Phillips, L. W. Short-term temporal changes in free recall. *Quart. J. Exper. Psychol.*, 1965, **17**, 132-138.
5. Raffel, G. Two determinants of the effect of primacy. *Amer. J. Psychol.*, 1936, **48**, 654-657.
6. Sumby, W. H. Word frequency and serial position effects. *J. Verb. Learn. & Verb. Behav.*, 1963, **1**, 443-450.
7. Thorndike, E. L., & Lorge, I. The Teacher's Wordbook of 30,000 Words. New York: Bur. Publ., Teachers Coll., Columbia Univ., 1944.
8. Tulving, E. Subjective organization in free recall of "unrelated" words. *Psychol. Rev.*, 1962, **69**, 344-354.
9. Tulving, E., & Arbuckle, T. Y. Sources of intratrial interference in immediate recall of paired associates. *J. Verb. Learn. & Verb. Behav.*, 1963, **1**, 321-334.
10. Waugh, N. C., & Norman, D. A. Primary memory. *Psychol. Rev.*, 1965, **72**, 89-104.
11. Welsh, G. B., & Burnett, C. T. Is primacy a factor in association formation? *Amer. J. Psychol.*, 1924, **35**, 396-401.

Department of Psychology
Vassar College
Poughkeepsie, New York 12601

Published as a separate and in *The Journal of Psychology*, 1968, 68, 243-248.

ATTEMPTED SUICIDE AS A HOSTILE ACT*[1]

Wellesley College

DAVID LESTER

A. INTRODUCTION

The purpose of this paper is to examine the implications of viewing suicidal behavior from the point of view of George Kelly's concepts of aggression and hostility. It will be pointed out that attempted suicide can be regarded as an act of hostility and that this statement includes many previous generalizations made about suicide. Kelly used the terms aggression and hostility differently from the traditional usage, and so it will be necessary first to examine analyses of suicidal behavior based upon the traditional meaning of these terms.

B. SUICIDE AS AN ACT OF INWARD-DIRECTED AGGRESSION

Following Freud (8), it has become a wide-spread assumption that suicide is an act of inward-directed aggression. For those psychoanalysts who accept the later dualistic position of Freud (9) and posit a life instinct and a death instinct, inward-directed aggression is primary, the direct derivative of the death instinct, whereas outward-directed aggression is a defense against primary self-destruction. For those psychoanalysts who remain monistic (and accept only a life instinct), aggression is seen as the result of frustration of needs in the individual and outward-directed aggression is often viewed as primary [see Futterman (10)].

The assumption that suicide can be regarded as an act of inward-directed aggression was also made by the sociologists Henry and Short (12), who considered suicide to be the opposite extreme of homicidal aggression. They argued that experience of psychological punishment (as opposed to physical punishment) favors the development of a strong super-ego and inhibition of aggression. When a child is psychologically punished and when the source of punishment and nurturance are the same, the child learns to inhibit his aggressive response to punishment, for this response threatens to end the supply of

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 8, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The ideas expressed in this paper come from a thesis submitted by the author to Brandeis University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

nurturance. They predicted, therefore, that suicidal individuals should have had more experience of psychological punishment than nonsuicidal individuals, a prediction based on the assumption that suicide is an act of inward-directed aggression.

Lester has attempted to test this prediction. In an investigation of a sample of nonliterate societies (19), no association was found between the discipline practices of the societies and the incidence of completed suicide in the societies. In a questionnaire study using college students (20), no association was found between recalled experience of punishment and threats or attempts at suicide.

Lester (16) argued that, if suicide is an act of inward-directed aggression, it should be related to body-image boundary dimensions, for Fisher and Cleveland (7) found that such dimensions were associated with the style of expressing aggression. In a study of psychiatric patients who had attempted suicide, Lester found no association between suicidal behavior and body-image boundary dimensions.

On the other hand, clinical studies have produced evidence that the suicidal individual is extremely hostile and aggressive. Offenkrantz *et al.* (23) found that delinquency, temper tantrums, an inability to postpone gratification, and poor work records characterized suicidal individuals. Yet psychoanalytic views see delinquency and suicidal behavior to be opposite kinds of disturbance and, therefore, find it difficult to explain their occurrence in the same individual. Braaten and Darling (3) found that suicidal college students were more psychopathic, as measured by the MMPI, than nonsuicidal students. They also noted that the suicidal students were more hostile, more angry, and showed excessive self-hate. West (25) has reported a high incidence of suicide in murderers.

Tabachnick (24) has described the existence of symbiotically dependent pairs of individuals, in whom suicidal behavior is common. He noted that one important facet of the relationship was the resentment felt by each member toward the other. Along these same lines is evidence reported by Lester (18, 21). He gave suicidal and nonsuicidal college students the Buss and Durkee hostility inventory (4). The two groups did not differ on the dimensions of assault, indirect hostility, verbal hostility, negativism, suspicion, or guilt. They differed only in that the suicidal individuals were more irritable and had more resentment than the nonsuicidal students. These results were found to be in accord with clinical reports; for example, Gould (11) noted the impulsivity and irritability of the adolescent suicide, and Balser and Masterson (2) have noted his excessive resentment. Lester concluded that there was no evidence that suicidal individuals possess a general trait of inhibiting aggression or turning it inwards onto themselves.

It appears, therefore, that there is little clinical or experimental support for the notion that the suicidal individual is less outwardly aggressive or more inwardly aggressive than the nonsuicidal individual. He is just as likely to be physically aggressive as the nonsuicidal individual. Suicidal ideation and psychopathic personality traits are often found in the same individual.

C. Suicide and Hostility

The notion that the individual who attempts suicide is in some way attempting to manipulate significant others in his environment is not new. Adler noted this (1) and the title of Farberow and Shneidman's book on attempted suicide emphasizes this—*The Cry for Help* (6). The suicidal individual indeed often does show evidence that he wishes to produce reactions indicative of love and attention from those close to him. Adler (1) noted the intent of the suicidal person to hurt others which in extreme cases casts suicide as an act of revenge. However, it is suggested here that this may not be the complete description of the suicidal person.

Many individuals are manipulators. The politician, the businessman, the salesman, and the poker player all manipulate others. Manipulation for some is an end in itself. The manipulative behavior of the suicidal individual does not appear to be of this kind. The suicidal individual is not Machiavellian.[2] Of what sort is the manipulative behavior of suicidal individuals? It is here that Kelly's notion of hostility is relevant.

Kelly (13, 14, 15) has written about those who manipulate others. For Kelly such individuals are hostile people. By "hostile" Kelly takes pains to show that he does not mean hostile in the usual sense of the word: namely, an impulse to hurt someone or to cause pain. Such a definition stresses the experience of the other person rather than the experience of the hostile person: "We consider you to be hostile when you hurt us." Kelly felt that hostility should be investigated from the point of view of the hostile person himself. How does it feel to the hostile individual to be hostile?

Kelly removed the notion of aggression from his definition of hostility. Aggression is one extreme of a passive-assertive dimension of behavior. Aggression is very similar to adventuresomeness. Hostility, on the other hand, lies in the reactions of the individual to information from the environment that is incongruent with his beliefs. When an individual meets such incongruent evidence he may accept it and attempt to erect a new system of beliefs from the ruins of the old system which will be consistent with the incongruent evidence. But he has an alternative. He can attempt to distort the data to fit his beliefs or he can attempt to extract alternative evidence from the environment that is congruent with his beliefs. Hostility then, for Kelly, is a refusal to accept and its opposite is realistic acceptance. The hostile person can be aggressive or passive but aggression is not an essential component of hostility.

---

[2] Lester in an unpublished study (17) found no association between attempted and threatened suicide in college students and scores on a Machiavellian scale devised by Richard Christie (5).

If the suicidal individual is to be considered a hostile person we must ask whether two necessary conditions are met. (*a*) Is he rigid in his thought and so unlikely to change or modify his beliefs? There is some evidence that this may be true. Neuringer (22), working from a different point of view, has reported that on a variety of psychological tasks suicidal individuals are more rigid in their thought than are nonsuicidal individuals or psychoneurotic patients. (*b*) To what end does the suicidal individual extort and distort evidence? The seeking of love and attention by the suicidal individual through his suicidal behavior has been noted above. The suicidal individual may expect love and attention from others and may believe that they do indeed love him. If they do not spontaneously give him evidence of this love, then his suicidal behavior may be an attempt to extort demonstrations of affection from them.

We have noted above that the suicidal individual is more resentful than the nonsuicidal individual. The suicidal person resents the world and feels that he has been treated unjustly. The suicidal person may act so as to give others a chance to reject him and so confirm his resentment against the world. By his suicidal behavior he risks being rejected by those who are important to him. The suicidal person makes it more difficult for his parents, his peers, and his wife or girl-friend to love and accept him. Often he is rejected after his suicidal gesture. Some suicidal persons may, therefore, be seeking rejection.

Suicidal behavior may serve to confirm other beliefs of the individual. The suicidal person frequently experiences feelings of worthlessness and depression. By risking rejection through his suicidal behavior, the suicidal individual can seek a confirmation of his worthlessness and a demonstration that he is not worth caring about.

D. CONCLUSION

The manipulative aspects of suicidal behavior have been examined. Mention has been made of suicide as an attempt to get love and attention from others and of suicide as an act of revenge. It has been suggested that the suicidal individual may also be seeking rejection through his suicidal behavior. Rejection serves to confirm his opinion of the world as an unfriendly and unjust place and also to confirm his opinion of his own worthlessness.

George Kelly's ideas on hostility and aggression were discussed and the behavior of the suicidal individual was interpreted as hostile behavior. This interpretation was considered to be more in accord with the experimental and clinical literature on suicide than the traditional psychoanalytic view of suicide as an act of inward-directed aggression.

REFERENCES

1. ANSBACHER, H. L. Suicide: The Adlerian point of view. In N. L. Farberow & E. S. Shneidman (Eds.), *The Cry for Help*. New York: McGraw-Hill, 1961.
2. BALSER, B. H., & MASTERSON, J. F. Suicide in adolescents. *Amer. J. Psychiat.*, 1959, **116**, 400-404.
3. BRAATEN, L. J., & DARLING, C. D. Suicidal tendencies among college students. *Psychiat. Quart.*, 1962, **36**, 665-692.
4. BUSS, A. H., & DURKEE, A. An inventory for assessing different kinds of hostility. *J. Consult. Psychol.*, 1957, **21**, 343-348.
5. CHRISTIE, R. Unpublished paper, Columbia University, New York, New York, 1967.
6. FARBEROW, N. L., & SHNEIDMAN, E. S. The Cry for Help. New York: McGraw-Hill, 1961.
7. FISHER, S., & CLEVELAND, S. E. Body Image and Personality. New York: Van Nostrand, 1958.
8. FREUD, S. Mourning and melancholia (1917). In *Collected Papers (Vol. 4)*. London: Hogarth Press, 1929.
9. ———. Beyond the Pleasure Principle. London: Hogarth Press, 1950.
10. FUTTERMAN, S. Suicide: Psychoanalytic point of view. In N. L. Farberow & E. S. Shneidman (Eds.), *The Cry for Help*. New York: McGraw-Hill, 1961.
11. GOULD, R. Suicide problems in children and adolescents. *Amer. J. Psychother.*, 1965, **19**, 228-246.
12. HENRY, A. F., & SHORT, J. F. Suicide and Homicide. Glencoe, Ill.: Free Press, 1951.
13. KELLY, G. A. The Psychology of Personal Constructs. New York: Norton, 1955.
14. ———. Hostility. Presidential address, Clinical Division, American Psychological Association, New York City, 1957.
15. ———. The threat of aggression. *J. Hum. Psychol.*, 1965, **5**, 195-201.
16. LESTER, D. Attempted suicide and body image. *J. of Psychol.*, 1967, **66**, 287-290.
17. ———. Unpublished paper, Brandeis University, Waltham, Massachusetts, 1967.
18. ———. Suicide as an aggressive act. *J. of Psychol.*, 1967, **66**, 47-50.
19. ———. Suicide, homicide, and the effects of socialization. *J. Personal. & Soc. Psychol.*, 1967, **5**, 466-468.
20. ———. Punishment experiences and suicidal preoccupation. *J. Genet. Psychol.*, in press.
21. ———. Suicide as an aggressive act: A replication with a control for neuroticism. *J. Gen. Psychol.*, in press.
22. NEURINGER, C. Rigid thinking in suicidal individuals. *J. Consult. Psychol.*, 1964, **28**, 54-58.

23. Offenkrantz, W., Church, E., & Elliott, R. Psychiatric management of suicide problems in military service. *Amer. J. Psychiat.*, 1957, **114**, 33-41.
24. Tabachnick, N. Interpersonal relations in suicide attempts: Some psychodynamic considerations and implications for treatment. *Arch. Gen. Psychiat.*, 1961, **4**, 16-21.
25. West, D. J. Murder Followed by Suicide. Cambridge, Mass.: Harvard Univ. Press, 1966.

Department of Psychology
Wellesley College
Wellesley, Massachusetts 02181

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 249-256.

THE ROLE OF THE FATHER IN GUIDANCE CLINIC CONSULTATIONS*[1]

Department of Psychology, The University of Calgary; and
The Provincial Guidance Clinic, Edmonton, Alberta

P. O. DAVIDSON AND A. R. SCHRAG

A. INTRODUCTION

The value of including the father as an integral part of any guidance clinic program has been increasingly emphasized since the early 1940's (3, 4, 13). While the importance of the mother-child relationship has certainly not been neglected (5, 12), increasing attention is now being directed to the family unit when considering both the aetiology (6, 7) and treatment (1) of emotional disturbance in children.

The difficulties encountered in engaging the cooperation and participation of the father in the treatment process have often been noted (9, 10, 11) and some of the causes for these difficulties systematically enumerated (2). One of the widely accepted ways of helping to overcome this resistance of the father to treatment is to get him involved at the earliest possible stage of treatment —usually the intake procedure (8).

Unfortunately a great many cases coming to the attention of guidance clinics never receive active treatment, and in some clinics a major portion of their time is spent in supplying diagnostic and consultative services rather than treatment. In this type of guidance clinic service, the role and importance of the father is even more uncertain than it is in the treatment procedure.

In providing rural service, for example, many guidance clinics can supply only very limited treatment facilities and a majority of their cases are seen only once or twice for diagnosis and recommendations to the family or community referring agency. Obviously, if this kind of limited service is to justify the considerable time and expense involved, the accuracy of the diagnosis and the carrying out of the recommendations are of crucial importance. It is equally obvious that the diagnosis and effectiveness of the recommendations are going to be influenced in part by the presence or absence of the father at the consultation. For this reason most clinics specifically ask that both parents attend when the child is first seen. Often, however, only one parent—usually the mother—brings the child to the clinic for diagnosis and consultation.

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 8, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was assisted by funds provided by the Mental Health Grant 608-7-66 of the National Health Grants Program. The authors are deeply indebted to Mr. B. Evans, Mrs. M. Francis, and Miss J. Smith for their invaluable assistance in collecting and analyzing the data for this study.

The present study was designed to assess some of the implications of whether both parents attend the initial contact with the guidance clinic. Specifically, answers were sought regarding the factors influencing the father's attendance at the guidance clinic and the effects of his presence or absence on the outcome of the consultation.

B. Method

The Edmonton Child Guidance Clinic has been providing, by regular travelling clinics, an extensive consultative service to the rural areas of northern Alberta. In addition to this rural service, many of the children in the city are also seen solely for diagnosis and consultation and not necessarily for treatment.

In 1962, for example, after excluding all treatment cases, 674 new rural cases had been seen for diagnosis or psychiatric consultation. Typically, in these cases, the child was tested by a psychologist, both parents (when available) were seen together by a social worker, and then the psychiatrist would see the child and parents—often together. In addition the case was usually discussed with the referring agency, and finally written reports and recommendations were sent to the appropriate persons involved in the case.

As part of a systematic review of the effectiveness of the Edmonton Guidance Clinic rural service, all the rural cases seen in 1962 were followed up—with particular attention directed toward differences found in relation to whether the father had attended the clinics in 1962. In addition a random sample of 150 of the 773 cases seen in the city of Edmonton for diagnosis only were selected for follow-up, to find out if there were any rural-urban differences in respect to the father's attendance at the guidance clinic.

This follow-up consisted essentially of retracing the 1962 travels of the Edmonton Guidance Clinic, arranging to see the parents, children, and referral sources that had been seen in 1962. Forms were sent out to the schools for any relevant information on changes in the child since 1962 and for his present behavior and school progress. Where parents were unable to attend the follow-up, specially prepared questionnaires were mailed to them at their current address or, whenever possible, arrangements were made to see them at a later date.

Out of the total of 824 cases from 1962, some follow-up information was obtained on 780 and complete information (including seeing at least one parent and the child) obtained on 518 of these cases. This represents direct follow-up on 63 per cent of all the cases, which is fairly typical for follow-up studies of this sort.

Both parents had attended the Guidance Clinic in 1962 in 31 per cent of the cases, in 51 per cent only the mother had attended, in 8 per cent of the cases only the father had attended, and in 10 per cent of the cases neither parent had attended. This last group included institutionalized children, children brought by welfare, probation officers, etc., or where parents were not available or refused to attend.

C. RESULTS

Since parental attendance in 1962 could be conveniently broken down into four discrete groups (father only, mother only, both parents, neither parent), chi squares were computed between this variable and relevant variables obtained from the follow-up of the cases.

Both parents attended significantly more frequently with children aged 4 to 10 and less frequently with children aged 12 to 18. For the age range 12—16 proportionately, more cases were seen with just the mother and, not surprisingly, for the age range 16—20 proportionately more cases were seen with neither parent ($\chi^2 = 168$, $p < .001$).

No significant differences were found between parental attendance and the sex of the child.

Rather surprisingly, in significantly more cases than expected both parents attended when the family was from a hamlet or farm and in significantly fewer cases both parents attended when the family was from a city or large town ($\chi^2 = 109$, $p < .001$). This may reflect the relative freedom of the farmer to leave his work for an hour or two compared to the urban father, but it may also reflect genuine rural-urban differences in the father's involvement in behavior problems in his children.

The father accompanied his child either alone or with the mother much more frequently when the guidance clinics were held at the local school. The mother only was seen proportionately more when the guidance clinic was scheduled in the local Public Health Office ($\chi^2 = 68$, $p < .001$). This finding was further reflected in the fact that proportionately greater attendance by the father occurred when the referral of the child was made by the school and proportionately more cases were seen with just the mother when the referral was made initially through the family doctor or public health nurse ($\chi^2 = 70$, $p < .001$).

A possible explanation for these results may be found in the parents' attitude toward their child's behavior. When parents are initially seen by the Guidance Clinic, one of the standard questions asked is about the parents' opinion of what they think is wrong with their child (if anything), and why they think the child was referred. It was found that the father was present either by himself or with the mother a significantly greater proportion of the time when the parental opinion was that the child was being seen for purely an educational problem ($\chi^2 = 58$, $p < .01$). Furthermore, the teacher's attitudes and the child's conduct in school also were found to be relevant to whether the father attended. When the child's school progress was poor and he was failing some subjects, the father accompanied him alone or with the mother significantly more often ($\chi^2 = 96$, $p < .001$). The father also appeared more often when the child's teacher was directly interested or involved in the child's current problem ($\chi^2 = 39$, $p < .01$).

On the other hand, just the mother accompanied the child more often when the child's conduct in the school was poor or disruptive and when the teacher's attitude toward the child was antagonistic or hostile. In addition when the mother came by herself she was significantly more likely to say the child was emotionally disturbed and not just an educational problem.

It was extremely interesting to note that while the father tended to be present when the parents thought there was an educational problem involved, and just the mother attended when the parents thought there was an emotional problem, this relationship did not hold when one considered the actual psychiatric diagnosis. A significantly greater proportion of both parents attended when the psychiatric diagnosis was emotional disturbance ($\chi^2 = 105$, $p < .001$) and no difference in parental attendance was found when the psychiatric diagnosis was one of strictly an educational problem. In other words it appeared as if the father became as involved as the mother in seeking help for a child that would be considered by a psychiatrist to be emotionally disturbed, but the father was much more apt to consider the problem in the light of the child's educational difficulties rather than his emotional adjustment. Just the mother seemed to be more concerned with the latter.

With the children diagnosed as psychotic, however, one of the most significant findings both statistically and clinically emerged. While slightly less than 2 per cent of all cases seen were diagnosed as psychotic, almost 10 per cent of all cases seen with just the father were psychotic. In other words, over four times as many as might be expected of the children who were diagnosed as psychotic were accompanied by just their father. As a matter of fact, 36 per cent of all the psychotic children were brought to the clinic accompanied by their father only. It is difficult to specify cause and effect in this relationship. Did the psychotic children have no mother to attend the clinic, was the mother too disinterested to attend, or was she unwilling or unable to attend? Inspection of the files of the psychotic children suggested that in many of the cases where just the father came, the mother was unwilling or unable to come because of what appeared to be serious mental disturbance (including alcoholism). Interestingly, no psychotic children were found in the group where neither parent attended.

Of the recommendations made for all the children seen in 1962, it was found on follow-up that in 35 per cent of the cases these recommendations were not carried out. This figure seems high in spite of the limited services and facilities available in many rural areas, but it is beyond the scope of this article to discuss reasons and problems associated with this result. Of more immediate relevance was the finding that when both parents were seen, proportionately more of the recommendations were fully or partly carried out, while in over 57 per cent of those cases where just the father was seen the recommendations were not carried out ($\chi^2 = 24$, $p < .01$). The only significant reason found for this failure to carry out recommendations was that the group where only the fathers attended disagreed more frequently with the psychiatric diagnosis than any other group ($\chi^2 = 32$, $p < .01$).

There were no significant differences between any of the groups in remembering what the recommendations had been, or in the proportion that felt the recommendations had been helpful.

Furthermore there were no differences in follow-up between any of the groups on their opinion of how much the child had changed.

There were significant differences, however, in terms of the follow-up team's evaluation of the present status of the child. It was found that significantly more of the children who were rated on follow-up as cured or improved had been seen in 1962 with their mother or both parents. Significantly more of the children who were accompanied just by their father were rated as unchanged and slightly more were rated as worse ($\chi^2 = 34$, $p < .001$).

There were no differences between any of the groups in the proportion changing their opinion on follow-up about what was wrong with their child. Of those who had changed their opinion following the guidance clinic consultation, there were no differences between any of the groups in the proportion who had changed their opinion toward closer agreement with the psychiatric diagnosis.

Finally it was noted that there were no differences between any of the groups in the proportion of children who were upset about going to the guidance clinic, although it was found that the child tended to be teased more by members of his family about going to the guidance clinic when just one parent had accompanied him (father or mother) than when he was accompanied by both.

D. DISCUSSION

From these results several tentative conclusions can be made. Firstly, it would appear that for rural child guidance consultations the presence of the father along with the mother may not be as essential as some psychiatrists have suggested. There were no significant differences between the cases where just the mother attended and where both parents attended in terms of whether the clinic recommendations were carried out, whether the child was improved on follow-up, or whether there had been any measurable change in parent's attitude to the child's behavior. When recommendations were not carried out there were no significant differences in the reasons for this failure between these two groups.

These findings would suggest that failure by the father to accompany the mother to the guidance clinic may reflect situational difficulties (inability to get away from work, etc.) or parental-role attitudes rather than any basic disinterest, negativism, or hostility toward the clinic. The apparent attitudes of the fathers about their role in seeking help for behavior problems with their children were fairly consistent for this study. It seemed that the fathers were quite willing to let the mothers assume responsibility in seeking help if the parents thought that the problem was primarily an emotional one and particularly one involving only social and interpersonal relations. This lack of involvement by the father was particularly evident in the children aged 12 to 20 referred to the guidance clinic for the first time. On the other hand, the father tended to be more involved and willing to seek help when it appeared to the parents that the child's difficulty was affecting his school work. This was particularly evident with younger children aged 4 to 10.

It was extremely encouraging to note, however, that when the child was actually emotionally disturbed (according to psychiatric diagnosis) the father was significantly more likely to accompany the mother to the guidance clinic, even though he claimed to be only interested in the related educational difficulties. In spite of this acknowledged interest only in the educational problems of the child, it turned out that the fathers were not particularly involved when the problem was solely an educational one (such as reading handicap, inadequate intelligence, etc.). In other words, it appeared that the father was often quite prepared to accompany the mother to seek help for an emotionally disturbed child but he often was not prepared to admit the underlying nature of the child's problem.

Secondly, these results focus attention on the complex dynamics and special problems involved in those cases which are brought to a guidance clinic consultation by the father alone. They would seem to constitute a special problem group for guidance clinic teams. The follow-up results of this group were rather discouraging. In a majority of these cases (and a significantly greater proportion than any other group) the recommendations were not carried out, the child's condition was unchanged or worse after two to three years, these parents disagreed with the psychiatric diagnosis, they saw only the educational problems and not the emotional problems of their children, and an alarming proportion of the children were found to be psychotic. It was surprising to find more difficulties associated with this group than with (as one might expect) the group where neither parent attended.

While these results may be unique to this sample, they certainly indicate that those child guidance cases accompanied by just the father deserve increased attention and further research.

Finally, this study points out some possible ways of increasing the involvement of the father in attending guidance clinics.

It would seem that the father is more willing to attend clinics that appear to be oriented toward the educational problems of the child: that is, clinics that are held in the local school, where the referral comes initially from the child's school teacher, and where the teacher is actively interested or involved in the case. In addition the father seems more inclined to accompany the mother when the parents can see the educational implications of the child's disturbed behavior.

E. SUMMARY

This study reported some of the results of a follow-up evaluation of 780 children who were seen by a guidance clinic team for diagnosis and psychiatric consultation. Emphasis was placed on the differences found in effectiveness of this kind of service as a function of which parent accompanied the child for the consultation. Chi square analysis of the follow-up results showed that the importance of having both parents' attendance at the clinic may be overemphasized—for this kind of guidance clinic service at least. Attention was drawn to the special problems and difficulties encountered in those cases that were accompanied by just the father, and the need for further research into the dynamics of this group was underlined.

REFERENCES

1. ACKERMAN, N. The Psychodynamics of Family Life. New York: Basic Books, 1959.
2. BELL, N., TRIESCHMAN, A., & VOGEL, E. A sociocultural analysis of the resistances of working-class fathers treated in a child psychiatric clinic. *Amer. J. Orthopsychiat.,* 1961, **31**, 388-405.
3. BERGMAN, M. The father gets worse: A child guidance problem. *Amer. J. Orthopsychiat.,* 1942, **12**, 474-486.
4. BERON, L. Fathers as clients of a child guidance clinic. *Smith Coll. Stud. Soc. Work,* 1944, **14**, 351-366.
5. GREEN, R. The treatment of parent-child relationships. *Amer. J. Orthopsychiat.,* 1948, **18**, 442-446.
6. HENRY, J. Family structure and the transmission of neurotic behavior. *Amer. J. Orthopsychiat.,* 1951, **21**, 800-818.
7. HENRY, J., & WARSON, S. Family structure and psychic development. *Amer. J. Orthopsychiat.,* 1951, **21**, 59-73.
8. MARTIN, F., & KNIGHT, J. Joint interviews as part of intake procedure in a child psychiatric clinic. *J. Child Psychol. & Psychiat.,* 1962, **3**, 27-39.
9. POLLACK, O. Integrating Sociological and Psychoanalytic Concepts. New York: Russell Sage Foundation, 1956.
10. POLLACK, O., *et al.* Social Science and Psychotherapy for Children. New York: Russell Sage Foundation, 1952.
11. RUBENSTEIN, B., & LEVITT, M. Some observations regarding the role of fathers in child psychotherapy. *Bull. Menninger Clinic,* 1957, **21**, 16-27.
12. SPERLING, M. The neurotic child and his mother. *Amer. J. Orthopsychiat.,* 1951, **21**, 351-364.
13. WALTON, E. Fathers as clients of a child guidance clinic. *Smith Coll. Stud. Soc. Work,* 1940, **11**, 155.

Department of Psychology
The University of Calgary
Calgary, Alberta, Canada

Published as a separate and in *The Journal of Psychology*, 1968, 68, 257-260.

PROGRAMMED INSTRUCTION, TEACHER ABILITY, AND SUBJECT MATTER DIFFICULTY*

Psychology Department, Northeast Louisiana State College

JOHN J. LA GAIPA[1]

A. INTRODUCTION

Numerous studies have compared programmed instruction with conventional instruction. Schramm (2) reviewed 36 comparative studies and found that 18 were inconclusive in finding no difference, 17 favored programmed instruction, and one favored conventional instruction. Stolorow (3) suggests that this type of research is premature because we know very little about the two complex phenomena we are comparing.

Most studies have failed to control for variations in teacher ability. Usually, there is no independent measure of the quality of conventional instruction; one does not know whether the specific comparisons dealt with good or poor teachers, and the sample of teachers is frequently inadequate, often consisting of only one teacher. Barcus, Hayman, and Johnson (1) found that teacher differences influence the effectiveness of programmed instruction. When teachers of Spanish classes were inexperienced, programmed instruction was superior to conventional instruction. This interaction was not replicated, however, in a second group of subjects that started the course at a different time.

The role of the teacher variable in comparative studies may be more evident if examined along with subject matter characteristics. The purpose of this study was to determine whether the effectiveness of programmed instruction, relative to conventional instruction, for teaching Basic Electricity might be influenced by two factors—ability of the teacher and subject matter difficulty.

B. METHOD

1. *Subjects*

The *S*s were 155 trainees from eight classes in the Basic Sonarmen Course, U.S. Navy Fleet Sonar School, Key West, Florida.

---

* Accepted by John A. Antoinetti of the Editorial Board; received in the Editorial Office, Provincetown, Massachusetts, on January 8, 1968; and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] A portion of this paper was read before the Southeastern Psychological Association, Atlanta, Georgia, April, 1967. The writer is indebted to Drs. Lloyd Standlee, Francis Harmon, and Mr. Eugene Hooprich for their valuable contributions to this study. The opinions and conclusions contained herein are those of the author and are not to be construed as representing those of the Department of the Navy.

2. *Teacher Ability*

Supervisors rated teachers on general instructional ability. The eight highest and the eight lowest rated instructors were used to teach the classes.

3. *Material*

Twenty instructors rated each unit of the Basic Electricity Course according to its complexity. The easiest unit was identified in the direct current part and the hardest unit was identified in the alternating current part of the course. Programmed booklets were prepared for these two units.

4. *Procedure*

The instructional methods consisted of programmed instruction and conventional instruction. Under the programmed method students worked on programmed booklets and the instructors answered questions but avoided lengthy explanations. Under the conventional method the usual lectures and demonstrations were given. To eliminate the confounding of experimental variables with any temporal effect, all conditions other than the DC-AC sequence were counterbalanced. This insured that each of the poor and good instructors taught both units under both methods. Parallel forms of objective achievement tests were administered. Covariance analysis was used to adjust achievement test scores in the eight classes to control for any differences in aptitude. Adjustments were made in terms of the Navy's General Classification, Mechanical, Clerical, and Reading Comprehension tests.

C. RESULTS

Table 1 presents adjusted mean test scores in the easy and hard units of the course. Separate analyses of variance were also conducted with covariance adjustments. For the *easy* subject matter, the method of instruction had a significant main effect on achievement ($F = 7.36$, $p < .01$ for 1,150 *df*). Programmed instruction yielded higher achievement scores than conventional instruction. Teacher ability also had a significant main effect on achievement ($F = 4.95$, $p < .05$ for 1,150 *df*). Good teachers yielded higher achievement than poor teachers. The interaction between method of instruction and teacher ability fell somewhat short of statistical significance ($F = 3.33$, $p > .05$ for 1,150 *df*).

Table 1 also presents data on student achievement in the *difficult* part of the course. The method of instruction again had a significant main effect ($F = 16.22$, $p < .001$ for 1,150 *df*). Teacher ability had no significant main effect

TABLE 1
COVARIANCE ADJUSTED MEAN TEST SCORES

Method of instruction	Teacher ability High	Low	Total
	Easy subject matter		
Conventional instruction	21.2 ($N = 41$)	18.8 ($N = 45$)	19.9 ($N = 86$)
Programmed instruction	21.7 ($N = 36$)	21.6 ($N = 33$)	21.7 ($N = 69$)
Total	21.4 ($N = 77$)	20.0 ($N = 78$)	20.7 ($N = 155$)
	Difficult subject matter		
Conventional instruction	12.1 ($N = 33$)	10.0 ($N = 36$)	11.0 ($N = 69$)
Programmed instruction	13.0 ($N = 45$)	13.9 ($N = 41$)	13.4 ($N = 86$)
Total	12.6 ($N = 78$)	12.1 ($N = 77$)	12.3 ($N = 155$)

($F = .55$, $p > .05$ for 1,150 *df*). A significant interaction, however, was found between method of instruction and teacher ability ($F = 6.72$, $p < .05$ for 1,150 *df*). Programmed instruction resulted in differentially higher achievement under poor teachers than under good teachers.

D. DISCUSSION

Examination of the data suggests that the degree of superiority of programmed instruction over conventional instruction is influenced by teacher ability and subject matter difficulty. The effects of programmed instruction apparently are greater with low ability teachers than with high ability teachers when the subject matter is complex. It appears that when the subject matter is easy, good teachers can handle it just as well without programmed instruction. When the subject matter is difficult, a poor teacher may do better with programmed instruction. Selective programming of the more complex material should be considered when there is a shortage of qualified instructors.

E. SUMMARY

Good and poor teachers taught both the easy and difficult units of Basic Electricity by programmed and conventional instruction. Overall, programmed instruction was more effective than conventional instruction. However, programmed instruction led to differentially higher achievement when the teachers were poor and the subject matter was difficult. The results of this study suggest

that, in order to make useful generalizations, comparative studies need to consider the quality of instruction and subject matter characteristics.

REFERENCES

1. BARCUS, D., HAYMAN, J. L., JR., & JOHNSON, J. T., JR. Programming instruction in elementary Spanish. In J. P. DeCecco (Ed.), *Educational Technology*. New York: Holt, Rinehart & Winston, 1964. Pp. 423-430.
2. SCHRAMM, W. The Research on Programmed Instruction: An Annotated Bibliography. Washington, D.C.: U.S. Office Educ., 1964.
3. STOLOROW, L. M. Implications of current research on future trends. *J. Educ. Res.*, 1962, **55**, 519-527.

Department of Psychology
Northeast Louisiana State College
Monroe, Louisiana 71201

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 261-266.

A NOTE ON THE USE OF JUDGES*[1,2]

Department of Psychology, Upsala College

LEWIS W. BRANDT

Methodological devices are introduced into scientific investigations and their rationale is rarely questioned thereafter. If the device has proven itself useful, it is frequently applied under other than the original circumstances without new justification on theoretical grounds. One such device in psychological experimentation is the use of "judges." They are commonly used for the selection of items, the creation of rating scales, and the assignment of responses to places on such scales. Neither the frequency with which "judges" are used nor the practical gains derived from their use can be regarded as justification for this method. It is the purpose of this paper to investigate the validity of the basic assumptions underlying the use of "judges," both "expert" and "nonexpert," by attempting to answer the following questions: (*a*) How can the use of judges be made congruent with accepted methodological standards in experimental psychology? And (*b*) how can the use of judges be justified on theoretical grounds? The writer will refer to the first as the methodological and to the second as the theoretical question.

A definition of the term "judge" as used in this context may be in order. Not every one who makes a judgment in a psychological experiment is referred to as a "judge." In various perception experiments, for example, we ask people for judgments but we call them "subjects." What then is the difference between "judges" and "subjects"? The writer's contention is that "subjects" are used to test a clearly defined, explicit hypothesis, while "judges" are used to support an undefined, implicit hypothesis. As an illustration, in one experiment, three groups of judges were used for the filtering of 100 "fitting" and 100 "nonfitting" words from the Thorndike-Lorge word count (4). The implicit hypothesis was that there are "fitting" and "nonfitting" words which can be recognized as such but not unequivocally defined in terms of their characteristics. This hypothesis was neither explicitly stated nor experimentally

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 8, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This paper was presented under the title "The use of judges in psychological experiments" at the 38th Annual Meeting of the Eastern Psychological Association, Boston, April, 1967.

[2] This paper was published with a grant from the Faculty Research Committee of Upsala College.

tested. Otherwise the "judges" would have been "subjects" in a preliminary experiment. Criteria for this hypothesis would have to be previously established and rejection of this first hypothesis would have led to the abandonment of the main experiment.

It could be asked at this point whether implicit hypotheses should not always be made explicit and experimentally tested. We shall return to this question in our theoretical considerations and use the above illustration presently only to illustrate the definition of "judges" as individuals used for the support of implicit hypotheses. With this definition we can now approach the methodological question concerning the use of "judges."

Experimental methodology requires the application of standard sampling procedures. Random samples are used to avoid bias in relation to the hypothesis and to allow for all possible contingencies. If subjects or stimuli are not randomly selected from the population or universe which they are supposed to represent, the results of the experiment could be due to the composition of the sample rather than to the conditions under investigation. A test of reliability becomes narrowly restricted in meaning, if subjects or items are selected in such a manner that a high correlation is built into the experiment due to the choice of the sample.

The same methodological considerations which apply to the selection of subjects or stimuli apply also to the selection of judges of criteria used in an experiment. If tests of reliability are applied to ratings, etc., a previously built-in reliability must be avoided lest the tests become meaningless through circularity. Good methodological procedure thus demands that the judges be chosen at random from a naive population: i.e., from among individuals who may be expected not to be able to figure out what the experimenter had in mind. Depending on the criteria which the judges are asked to establish, they may be selected so as to represent both sexes and various educational, socioeconomic, racial, religious, etc., backgrounds.

These sampling criteria are disregarded when so-called "expert judges" are used. They usually are persons with homogeneous backgrounds, views, and prejudices. The frequently encountered high correlation between judgments of such experts may be a necessary consequence of the way in which the judges were selected. Conclusions drawn by the experimenter concerning the validity of criteria based on high correlations between expert judgments are thus fallacious. The very bias which the random sampling of subjects or stimuli attempts to reduce is being introduced into the experiment through the use of judges selected not randomly.

A second methodological issue concerns the use of introspection in an experiment which investigates observable behavior. In such an experiment introspection is commonly ruled out as an experimental method. Judgments by raters are, however, usually based on introspection and not on objectifiable data. When an experiment uses instruments which are built upon such introspectively arrived at judgments, the latter needs must pull the rug out from under the claim for objectivity concerning the experimental findings, unless those findings are *reported* as opinions, judgments, beliefs, etc.

The second issue to be raised in this paper pertains to the theoretical assumptions underlying the use of judges in designing procedures or instruments for an experiment. There are a number of institutions in our society in which judges are used: e.g., the juries in courts of justice and beauty contests and the umpires in sports. Since these are not sciences, no justification of the use of judges is required on theoretical grounds. What all of these judges have in common is that they are asked to decide whether and to what extent the rules of the game have been abided by, or they have to make judgments of taste. They are not asked to establish criteria but to apply pre-established criteria. Most important, their opinions are accepted as opinions and not as scientific truth.

The issue is quite different in an experimental study. Such a study is set up to discover scientific truth. Opinions must, therefore, be completely eliminated from experimental research. If the instruments used in an experiment are based on opinions, the findings must contain these opinions as a factor. There can be no further claim for objectivity in the final data except that it may be objective to conclude that "so many were for and so many against . . . ."

In order to decide whether or not the judges are asked to give opinions, we must inquire into the kind of decisions they are asked to make. Looking then at the kind of judgments made in establishing rating scales, selecting stimuli, etc., we find that judges are required to determine relationships. Basically, these are relationships between scale items and certain aspects of behavior or experience or between experimental procedures and theoretical concepts.

When we are dealing with objectifiable data, such as scale items and behavior, relationships can be established only by experimentation. A specific relationship exists or not, and cannot be established by the opinion of judges.

More frequently judges are called upon to establish relationships between test items or other stimuli and theoretical concepts. In other words, the judges are really asked to confirm the experimenter's definitions. In a disguised way, this is the case in most ratings by so-called "experts." The logical error lies here in the assumption that a definition can be supported by experts.

tested. Otherwise the "judges" would have been "subjects" in a preliminary experiment. Criteria for this hypothesis would have to be previously established and rejection of this first hypothesis would have led to the abandonment of the main experiment.

It could be asked at this point whether implicit hypotheses should not always be made explicit and experimentally tested. We shall return to this question in our theoretical considerations and use the above illustration presently only to illustrate the definition of "judges" as individuals used for the support of implicit hypotheses. With this definition we can now approach the methodological question concerning the use of "judges."

Experimental methodology requires the application of standard sampling procedures. Random samples are used to avoid bias in relation to the hypothesis and to allow for all possible contingencies. If subjects or stimuli are not randomly selected from the population or universe which they are supposed to represent, the results of the experiment could be due to the composition of the sample rather than to the conditions under investigation. A test of reliability becomes narrowly restricted in meaning, if subjects or items are selected in such a manner that a high correlation is built into the experiment due to the choice of the sample.

The same methodological considerations which apply to the selection of subjects or stimuli apply also to the selection of judges of criteria used in an experiment. If tests of reliability are applied to ratings, etc., a previously built-in reliability must be avoided lest the tests become meaningless through circularity. Good methodological procedure thus demands that the judges be chosen at random from a naive population: i.e., from among individuals who may be expected not to be able to figure out what the experimenter had in mind. Depending on the criteria which the judges are asked to establish, they may be selected so as to represent both sexes and various educational, socioeconomic, racial, religious, etc., backgrounds.

These sampling criteria are disregarded when so-called "expert judges" are used. They usually are persons with homogeneous backgrounds, views, and prejudices. The frequently encountered high correlation between judgments of such experts may be a necessary consequence of the way in which the judges were selected. Conclusions drawn by the experimenter concerning the validity of criteria based on high correlations between expert judgments are thus fallacious. The very bias which the random sampling of subjects or stimuli attempts to reduce is being introduced into the experiment through the use of judges selected not randomly.

A second methodological issue concerns the use of introspection in an experiment which investigates observable behavior. In such an experiment introspection is commonly ruled out as an experimental method. Judgments by raters are, however, usually based on introspection and not on objectifiable data. When an experiment uses instruments which are built upon such introspectively arrived at judgments, the latter needs must pull the rug out from under the claim for objectivity concerning the experimental findings, unless those findings are *reported* as opinions, judgments, beliefs, etc.

The second issue to be raised in this paper pertains to the theoretical assumptions underlying the use of judges in designing procedures or instruments for an experiment. There are a number of institutions in our society in which judges are used: e.g., the juries in courts of justice and beauty contests and the umpires in sports. Since these are not sciences, no justification of the use of judges is required on theoretical grounds. What all of these judges have in common is that they are asked to decide whether and to what extent the rules of the game have been abided by, or they have to make judgments of taste. They are not asked to establish criteria but to apply pre-established criteria. Most important, their opinions are accepted as opinions and not as scientific truth.

The issue is quite different in an experimental study. Such a study is set up to discover scientific truth. Opinions must, therefore, be completely eliminated from experimental research. If the instruments used in an experiment are based on opinions, the findings must contain these opinions as a factor. There can be no further claim for objectivity in the final data except that it may be objective to conclude that "so many were for and so many against . . . ."

In order to decide whether or not the judges are asked to give opinions, we must inquire into the kind of decisions they are asked to make. Looking then at the kind of judgments made in establishing rating scales, selecting stimuli, etc., we find that judges are required to determine relationships. Basically, these are relationships between scale items and certain aspects of behavior or experience or between experimental procedures and theoretical concepts.

When we are dealing with objectifiable data, such as scale items and behavior, relationships can be established only by experimentation. A specific relationship exists or not, and cannot be established by the opinion of judges.

More frequently judges are called upon to establish relationships between test items or other stimuli and theoretical concepts. In other words, the judges are really asked to confirm the experimenter's definitions. In a disguised way, this is the case in most ratings by so-called "experts." The logical error lies here in the assumption that a definition can be supported by experts.

Definitions are, however, based on conceptualizations or came into being by a social process and can, therefore, be neither supported nor rejected. They cannot be true or false. Definitions can be only more or less clear or ambiguous.

What is being tested in a reliability test which uses the correlation of judgments by several "experts" is then the extent to which the judges have understood the experimenter's definitions in the way in which he meant them. Thus, judges are used to test the communicability of definitions.

Seen in this light, the use of judges and even of expert judges can be justified on theoretical grounds. Since the experimenter intends to communicate his definitions to others, he may want to test the communicability of his concepts. If his prospective audience consists exclusively of certain experts, the experimenter is justified in testing the communicability of his concepts only on experts. However, several conditions ought to be fulfilled in this instance: (*a*) The judges must form a representative sample of the population to whom the experimenter wishes to communicate his research. (*b*) Terms cannot be pooled but interjudge agreement must be obtained independently for each concept (scale item, rating, response, etc.). (*c*) Unless interjudge agreement reaches a previously established level of confidence, the concept must be redefined. It ought to be clear that what is being measured by the correlation of interjudge agreement is the clarity of a definition: i.e., the relationship between a sign and its referent and not the relationship between two observable events.

Since this is an issue of communication, one might raise the question as to whether rating procedures are the most efficient and reliable way to determine the clarity of one's definitions. One might decide that the clarity of a definition can be tested more validly by giving it to someone with the instructions to list specific items which meet the definition. If the other person names only items which in the conceptualization of the definer fall within the realm of his definition, the latter may be considered sufficiently free from ambiguity. This procedure might even be considered to meet the requirements of a psychological experiment.

The writer would like to illustrate the above considerations on the basis of two studies in which judges were used. In the first one, (4) judges were required to judge the "fittingness" of word-sound to meaning. The author gives no definition of "fittingness." He selected from the high-frequency categories of the Thorndike-Lorge word count 100 "fitting" and 100 "nonfitting" words. Three groups of judges served as filters for this list. Five "fitting" and six "nonfitting" words were obtained on which all judges agreed. Ten of these words were retained for the subsequent experiments. However, of the 24 *S*s originally chosen for the experiments, nine *S*s had to be excluded because they did not rate all 10 words in agreement with the original judges in terms of "fittingness." *E*'s implicit assumption was that the judges established a relationship between a sign and a referent: namely, between the term "fittingness" and certain words which "have" this characteristic. He did, however, nothing of the sort, as evidenced by the one-third of the *S*s whom he had to exclude because the relationship did not exist for them. Had *E* clearly defined his concept of "fittingness" by what Holzkamp (2) calls a "phenomen-analytical circumscription" and tested the communicability of this definition on some of his colleagues, he would have been on firm theoretical grounds for the experimental usage of members of the class of "fitting words."

As a second example the writer refers to his own doctoral dissertation (1) in which his use of judges was never questioned by any member of his Doctoral committee. The writer had hypothesized that applicants for psychotherapy who do not show up even for their first interview present their problems in a written application for treatment differently from those who eventually come for therapy. He had established eight criteria on which the statements of problems were hypothesized to differ. Three criteria concerned the content of the written statement; one was concerned with length and four were linguistic. Two judges who were unfamiliar with the purpose of the study and the author himself classified the statements. Interjudge agreement was far beyond chance. Both the writer and his committee implicitly assumed that the judges established a relationship between the eight categories and the applicants' statements of problems. Had the writer been able to define such categories as "patient presents his problem as lying within himself or in his environment" and "use of technical psychological terms" with sufficient clarity, he could have established the required relationship by means of definitions. In the absence of such a clear definition, no number of judges could establish that relationship. The relationship between a sign and its referent or between a concept and its definition is a *con*ceptual one. It is not a *per*ceived relationship. Thus, there is nothing to be "validated" by judges.

It might be argued that many theoretical concepts cannot be unambiguously defined because they fall into the realm of inarticulate knowledge (3). If this point of view is accepted, it seems to follow that we are dealing with "personal knowledge" and that experimental research becomes largely questionable as to its significance for science.

SUMMARY

The methodological and theoretical implications of the use of judges are discussed and illustrated by two examples. It was concluded that judges cannot

replace definitions or validate relationships but can only establish the communicability of concepts. If used for this purpose, they ought to be randomly selected from the population with which the experimenter intends to communicate.

REFERENCES

1. BRANDT, L. W. Psycholinguistic analysis of statements of problems from applicants for psychotherapy: A comparison of a group who rejected and a group who accepted treatment at three outpatient mental health clinics. Ph.D. dissertation, New York University, New York, 1963.
2. HOLZKAMP, K. Theorie und Experiment in der Psychologie. Berlin: de Gruyter, 1964.
3. POLANYI, M. Personal Knowledge. New York: Harper & Row, 1958.
4. WERTHEIMER, M. The relation between the sound of a word and its meaning. *Amer. J. Psychol.*, 1958, **71**, 412-415.

Department of Psychology
Upsala College
East Orange, New Jersey 07019

Published as a separate and in *The Journal of Psychology*, 1968, 68, 267-280.

A PSYCHOLOGICAL ANALYSIS OF "RESISTANCE" FIGHTERS DURING WORLD WAR II*

Bryn Mawr, Pennsylvania

ANDRE L. JULIARD AND ALEX S. JULIARD

A. INTRODUCTION

A new insight into the nature of human motivation and into our own psychological makeup may sometimes result from the observation of individuals living in unusual conditions. These conditions occurred, for example, among people who joined the French "resistance" during World War II. At that time the authors had the opportunity to observe the behavior of young and middle-aged persons who belonged to a "maquis" in a mountainous region of Southeast France. Their observations induced them to discern in human beings a larger variety of innate aptitudes, or inclinations, than those currently recognized in normal daily existence. Some of these overlooked inclinations, nevertheless, play an important part in the behavior of dedicated people.

This essay describes the nature of these inclinations and establishes a matrix which presents a graduated scale of words designating basic behavior patterns, or elemental "attitudes." This matrix gives an unequivocal meaning to the terms of the unprecise vocabulary used in daily language to describe the temperament of a person. The character matrix is used in this paper to establish the personality profile of different types of "resistance" fighters.

It is currently accepted that human activity is the result of various desires or fears determined by biological and social drives. The intensity of these drives varies from one individual to the other, depending more or less upon the state of development of one's psychological inclinations. However, there is little agreement among psychologists about the nature of these inclinations, since their number varies, depending on the psychological school, from three (2) to 40 (3).

The authors' personality study was begun with the application of the Delmas and Boll (1) point of view, which explains human behavior by the interplay of five affective and three intellectual inclinations. The system of personality analysis of Delmas and Boll was the result of the analysis of emotional disturbances observed among shell-shocked soldiers during World War I. By

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 8, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

comparing the behavior of some of their patients with that of normal persons, they concluded that the personality of an individual is determined by the degree of development of the five following affective dispositions: "avidity," the desire to acquire; "sociability," the desire to be appreciated; "kindness," the desire to be helpful; "activity," the desire to be continuously busy; and "emotivity," the ability to be emotionally sensitive. Well defined psychological derangements occur when one or more of these dispositions becomes under- or overdeveloped. The basic behavior patterns, or elemental "attitudes," which correspond to the developments just short of the pathological state of these dispositions are given in Figure 1. The intensity of the inclination corresponding to these extreme attitudes was arbitrarily scaled by Delmas and Boll —3 and +3, respectively.

DISPOSITIONS	EXTREME PERSONALITY TRAITS	
	Underdeveloped(-3)	Overdeveloped(+3)
Avidity	Disinterested	Acquisitive
Sociability	Cynic	'Feigner'
Kindness	Mean	Compliant
Activity	Depressed	Excited
Emotivity	Impassive	'Sensitive'

FIGURE 1

CLASSIFICATION OF ELEMENTAL ATTITUDES[a]

[a] According to Delmas and Boll (1).

By observing the behavior of, and talking casually with, people unaware of their goal, the authors soon realized that man's conduct was the result of a greater number of motivations than those acknowledged by Delmas and Boll. While some of these motivations, like the fear of being drafted as a worker in Germany or the desire to be considered a hero, had a biological or a social origin, others, like the desire for revenge or a search for self-appreciation, appeared to stem from other drives.

After a long period of groping and more than three years of trial and error, the authors came to differentiate 20 independent inclinations and for each of them, four elemental attitudes. These inclinations and their corresponding attitudes are shown in Table 1.

The following comments will help in understanding how this table was

TABLE 1

Classification of Elemental Attitudes

Nature of drive	Inclination	Basic personality traits			
Intensity of inclination		0	1	2	3
Psychic	Intuitiveness	“Spiritless”	Semiintuitive	Intuitive	Psychic
	Aestheticism	“Crude”	“Tasteless”	“Refined”	Aesthetic
	Self-appreciativeness	Self-deprecating	“Unsure”	Self-accepting	Smug
Volitional (Will)	Concentrativeness	Scatterbrained	Distractable	Attentive	Concentrated
	Imperturbableness	Impulsive	Emotional	Self-controlled	Imperturbable
	Firmness	Weak-willed	Irresolute	Perseverant	Obstinate
Mental	Imaginativeness	“Formalist”	Conventional	Imaginative	“Original”
	Conceptiveness	Illogical	Reasonable	Logical	Conceptive
	Perceptiveness	Unperceptive	“Aware”	Perceptive	“Keen”
Ethical	Forgiveness	Vengeful	Resentful	“Unresentful”	Forgiving
	Compassion	Heartless	“Unconcerned”	Humane	Compassionate
	Righteousness	Unjust	Partial	Just	Upright
	Sincerity	False	Insincere	Sincere	Outspoken
Intensity of inclination		3	2	1	0
Social	Domineeringness	Domineering	Uncooperative	Cooperative	Submissive
	Vanity	“Show-off”	“Vain”	“Natural”	Self-effacing
	Gregariousness	Gregarious	Sociable	“Discriminating”	“Distant”
Biological	Lasciviousness	Lascivious	Sensual	“Decent”	“Frigid”
	Acquisitiveness	Acquisitive	Possessive	“Provident”	Improvident
	Gourmandism	“Gluttonous”	“Gourmand”	“Temperate”	“Abstemious”
	Fearfulness	“Fearful”	Cautious	Brave	Fearless

built and how it was applied to analyze the personality profile of resistance fighters.

1. This classification of attitudes differs from present classical personality analysis systems in that it separates ethical and psychical drives from biological and social ones. It is the opinion of the authors that some attitudes, such as compassion or forgiveness, are the result of motivations which transcend the "commands" of parental conscience and of the rules of society. The authors assume also that psychic sensitivity, like the ability to be enthralled by a sunset, as well as the possibility of experiencing strange awarenesses through deep meditation, cannot be ascribed to biological or social needs.

2. The degree of development of an inclination can be evaluated by the effort and even the sacrifice one is inclined to make either to reach its gratification, which can be pleasure, joy, or happiness; or to avoid its suffering, which can be pain, sorrow, or sadness. The authors tried to establish four stages in the degree of development of each inclination, and tried, furthermore, to identify each corresponding elemental attitude, or basic personality trait, with a name from the usual vocabulary.

The identification of each elemental attitude is complicated by the normal interplay of the desires and fears pertaining to the different inclinations. For example, a person naturally inclined to judge objectively may act unjustly for many motives. He may commit an injustice to satisfy some personal interest if he is greedy. He may also act unjustly to yield either to the pressure of his superior if he is submissive or to the threat of a gang if cautious or fearful.

3. The correct word to designate a specific attitude was difficult to find because the classical personality traits often correspond to compound attitudes. It was particularly difficult to find words expressing attitudes in which no act of will[1] interferes. "Courage," for example, implies an interference of the will power when one forces oneself to face a danger. This is the reason for which the word "brave" is used—and not "courageous" as was used by Aristotle—to designate the genuine elementary "virtuous" attitude some people show toward the threat of physical suffering.

The choice of the terms used in Table 1 was determined by logic and semantics. Each of the terms which designates the four degrees of development of the same disposition must be incompatible with the three other terms, for a person cannot be at the same time a show-off and self-effacing. On the other hand, each of these four terms must be compatible with all the other terms of the table if we assume the unrelatedness of the inclinations. For example, a show-off may be at the same time either sincere or false, forgiving or resentful, and so forth.

---

[1] The nature of this controversial psychological activity will be discussed below.

An adequate word could not always be found to designate some elemental attitudes because our language did not evolve toward such a goal. When this happened, the word the authors chose was the one whose meaning in one of his usages came the closest to describing the basic personality trait framed by the classification. For instance, there are no adequate English terms to designate two states of sexual attitude below "sensual," which already scores high in sexual inclination. *"Decent"* and *"frigid"* were chosen: *"frigid"* is used in its vernacular meaning instead of its medical one, while *"decent"* is used with the connotation of "being decently responsive to sexual desires." These words are written in italics or between quotation marks to emphasize the slight stretching in their linguistic use when applied to designate a basic personality trait.

4. The number of possible different personalities corresponding to the four modalities of the 20 inclinations is equal to 4^{20} or approximately one trillion, a number which represents 250 times the actual earth population. Fortunately, a personality can be characterized only through the few aptitudes which correspond to inclinations that are either insufficiently or excessively developed, in the same way that a face can be sketched with its few main characteristic lines.

5. Each specific elemental attitude depends both on inborn imprints and on acquired training, through education and self-discipline. Sometime an abrupt mutation in fundamental attitude occurs. This happens usually when an individual is subjected to a tragic experience, like the loss of a beloved one or the realization that ethical and psychic factors may intervene in the happenings of his life.

We shall now proceed with the description of the inclinations and the related fundamental attitudes given in Table 1.

B. BIOLOGICAL DRIVES

One can distinguish four fundamental biological driving forces aimed at protecting the life of the individual and ensuring his offspring. The biological inclinations are here named by the fault corresponding to the exaggerated form they sometime display: "fearfulness," "acquisitiveness," "lasciviousness," and "gourmandism."

Fearfulness is the most essential biological inclination, for fear of pain and injuries generates reactions which usually protect a living organism from destruction. The sequence of attitudes corresponding to increasing instinctive fearfulness is "fearless," "brave," "cautious," and *"fearful."* A fearless person, like President J. F. Kennedy, takes unnecessary risk for himself, while a "fearful" one is so disturbed by fear that he avoids the reasonable risk one must take to act as a reliable operator, a loyal citizen, or a successful creator.

Gourmandism is the inclination which pushes a living organism to sustain its metabolism through feeding. The intensity of this inclination can be evaluated by observing the importance good eating and drinking plays in the behavior of a person. The sequence of attitudes corresponding to increased gourmandism is *"abstemious" "temperate," "gourmand,"* and "gluttonous." In this classification, a "gourmand" is a person who finds a great pleasure in eating and drinking, while a "glutton" craves that pleasure. "Gourmet" is a compound attitude designating the behavior of a "refined" "gourmand."

Acquisitiveness is the inclination corresponding to the drive which pushes a living organism to accumulate provisions in foresight of scarcity. The acquisitiveness of a person can be evaluated by the intensity of his drive to acquire and keep material goods. Persons with increasing strength in this inclination would be considered as "improvident," *"provident,"* "possessive," and acquisitive, respectively. An acquisitive person enjoys the accumulation of wealth solely for the pleasure of accumulation. The extreme attitudes, improvident and acquisitive, are both failings, for the former may jeopardize security and the latter the improvement of the individual.

Finally, lasciviousness designates the sexual inclination which ensures the continuity of living species. One is either *"frigid," "decent,"* "sensual," or "lascivious," depending on the degree to which this inclination is displayed. The behavior of a "frigid" person is—without the pressure of moral or religious constraint—not governed by sexual desire, while a "decent" person pays only little attention to sex in daily activity. A lascivious person is mainly preoccupied by sexual matters and usually sacrifices all other considerations to reach erotic gratification.

C. Social Drives

The social inclinations are the source of the drives which favor the formation of flocks, and also play a part in the formation of more specifically human groups like clubs, political parties, and nations. There seem to be only three social inclinations which, like the biological ones, are designated by the attitude corresponding to the exaggerated state of their development: "gregariousness," "vanity," and "domineeringness."

"Gregariousness" is the word that describes the inclination of people to group together for the simple pleasure of togetherness. Following the increase in strength of this inclination, one is *"distant," "discriminating,"* "sociable," or "gregarious." A "distant" person does not need company, and likes to be alone without being therefore necessarily a misanthrope. A gregarious person craves the society of other people and sometime makes serious effort to attract their affection.

"Vanity" designates the inclination which pushes an individual to attract the attention of other people. The primitive usefulness of this drive is obviously the improvement of the individual and, consequently, that of the group by fostering emulation among its members. The elemental attitudes corresponding to an increase in the strength of vanity are "self-effacing," *"natural," "vain,"* and *"show-off."* This inclination is not necessarily associated with self-appreciation, for one can be smug—i.e., show an exaggerated good opinion about one's self—and, nevertheless, remain self-effacing. This is the reason for which *"natural"* was chosen for the third step in the vanity scale instead of "modest," which connotation implies an element of self-appreciation.

"Domineeringness" designates the inclination shown by some people to control and even to rule the behavior of others. This inclination manifests its utility by favoring the establishment of a hierarchy, and therefore a corporated structure, into the group. The sequence of elemental attitudes corresponding to an increase in domineeringness is "submissive," "cooperative," "uncooperative," and "domineering." A submissive person completely abdicates his independence of action, while a domineering one tries to impose his point of view on others. The latter will succeed if he has, furthermore, strong mental and volitional aptitudes. In this case, such a domineering person will act, depending on the development of his ethical inclinations, either as a tyrant or as a leader.

D. Ethical Drives

As already stated, the reality of ethical drives remains controversial because ethical drives are seldom as strongly developed as the biological or the social ones. Nevertheless, the behavior of many outstanding personalities, like Baruch Spinoza, Mahatma Gandhi, or Albert Schweitzer, remains unintelligible unless we entertain the assumption that they were motivated by higher goals than their own interest or the interest of the social or ethnical group to which they belonged. It seems that the so-called ethical motivations are rooted in a feeling of reverence toward all human beings, one's self included as expressed in self-respect. Tragic dilemmas may arise from conflict between biological desire or social pressure and the sense of responsibility dictated by ethical feeling.

There seem to be four ethical, independent inclinations: "sincerity," "righteousness," "compassion," and "forgiveness."

"Sincerity" refers to the inclination to express one's self in accordance with what one considers to be factually or logically true. The sequence of elemental attitudes corresponding to an increased urge to tell the truth is "false," "insincere," "sincere," and "outspoken." A false person does not have any scruples in telling a lie, while a sincere person will try his best to avoid such saying. To be outspoken, which means to give unsolicited sincere opinions, could be detrimental to one's self or to others.

"Righteousness" designates the inclination to defend the civil and moral rights of others. The intensity of this inclination can be evaluated by the effort one develops to defend what one considers to be right. The elemental attitudes corresponding to an increase of this inclination are "unjust," "partial," "just," and "upright." A just person is known for the fairness of his action, while an upright person shows an outraged sense of justice or morality and tries to restore them at any cost. An unjust person is not perturbed in the least when justice is a mockery.

"Compassion" designates the inclination to be moved by the suffering and sorrow of others. Following the increase of the urge to alleviate the suffering of unrelated persons, one is "heartless," *"unconcerned,"* "humane," or "compassionate." While a heartless person is indifferent to any suffering, even that which he causes, an "unconcerned" person is oblivious to the pain of others. A humane person tries to help persons in need within the limits of his possibilities, while a compassionate person is so strongly moved by compassion that he will help others at the price of a serious sacrifice.

"Forgiveness" expresses the inclination of an individual to pardon offenses and to react without any resentment toward the person or the group which caused him a physical or moral damage. The sequence of basic behavior patterns corresponding to an increased ability to forgive is "vengeful," "resentful," *"unresentful,"* and *"forgiving."* According to this classification a "forgiving" person not only feels no resentment against the offender, but forgets completely the endured harm, overlooking any possibility of recidivism.

E. Mental Drives

Mental activity is a computer-type function of the brain which helps the individual to reach the goals "suggested" by his biological, social, and ethical drives. This activity seems to be related to three independent inclinations: "perceptiveness," "conceptiveness," and "imaginativeness."

"Perceptiveness" designates the inclination to identify objects or situations by comparing coordinated sensory impressions with similar ones stored in the memory. Visual acuteness plays an important part in perceptiveness, which is usually more developed among "doers" than among "thinkers." A person is said to be "unperceptive," *"aware,"* "perceptive," or *"keen,"* according to the degree of development of this aptitude. An unperceptive person will be blurred by lack of percepted information when placed in an unusual surrounding. A "keen" person placed in the same environment will be overwhelmed by the abundance of simultaneously percepted information.

"Conceptiveness" designates the aptitude of discerning similarities between different perceptions and of expressing these similarities by symbols. At a higher level, abstract operations can be applied between concepts themselves. Following the increase in his ability to abstract, a person will be considered as *"illogical," "reasonable,"* "logical," or "conceptive." An illogical person shows no propensity to speculate. A person strongly motivated by scientific curiosity usually scores high in abstractiveness.

"Imaginativeness" qualifies the mental inclination to play with memories. Through this aptitude one is able to fragmentize memorized perceptions, especially visual ones, and to reassemble the pieces into new, original patterns. In this mental dimension a person will be considered either as *"formal,"* "conventional," "imaginative," or *"original."* A person must be "original," or at least imaginative, to be able to become a creative scientist or a talented artist.

F. Volitional Drives

The power of action of an individual upon the outside physical world and upon his inside emotional and mental world is related to the apparently obsolete concept of will. There seem to be three independent types of will activities: "firmness," *"imperturbableness,"* and "concentrativeness."

"Firmness" refers to the inclination to influence and control the behavior of physical entities, human beings included. A person is considered as "weak-willed," "irresolute," "perseverant," or "obstinate" according to the degree of development of his firmness. For a perseverant or an obstinate person, the quest for the satisfaction which accompanies the achievement of a task becomes a powerful driving force and constitutes an important component in its psychological makeup.

"Imperturbableness" designates the inclination of controlling feelings and emotions. With the increase in "imperturbableness," a person will be categorized as "impulsive," "emotive," "self-controlled," and "imperturbable." Here again, while the third term of this sequence of attitudes, to be self-controlled, is a desirable quality, the last one, to be imperturbable, can be considered as a fault.

"Concentrativeness" is the aptitude to direct one's attention for a long interval of time on a single object or on a single thought. The sequence of elemental attitudes which corresponds to increasing ability in concentration is "scatterbrained," "distractable," "attentive," and "concentrated." "Distractable" is used in the sense of being easily distracted from the object of one's attention, and not of being inattentive as the result of other preoccupations. A concentrated mind can be so deeply absorbed in its meditation that awareness of the outside world may disappear during this process.

G. Psychic Drives

We are now reaching the most controversial part of this study, which deals with concepts that are hardly operationally definable in spite of the fact that they are related to phenomena which act on some people as powerful driving forces. The word "psychic" covers this kind of knowledge. It deals with deeper types of awareness than being conscient of sense perception, emotional feelings, or thought manipulation. Among these awarenesses is the appreciation of one's self-accomplishments, or "self-appreciativeness"; of harmony in nature, or "aesthetism"; and the comprehension of what is called "spiritual realities," or "intuitiveness."

"Self-appreciativeness" designates the inclination to be more or less satisfied with one's own behavior and accomplishments. With increasing satisfaction, one is either "self-deprecating," "unsure," "self-accepting," or "smug." The attitude of a person in this dimension is the result of his life's experience.

"Aestheticism" expresses the hard-to-explain motivation which pushes some people to search for "harmonious" impressions and expression. This harmony can be displayed in form, color, sound, in an artistic creation, a natural scenery, or displayed in the logical arrangement of a mathematical or a scientific work. The fundamental attitudes corresponding to increased aestheticism are *"crude,"* "tasteless," *"refined,"* and "aesthetic." While a "crude" person is completely insensitive to aesthetic pleasure, an aesthete is so enthralled in his quest for beauty that other possible accomplishments are sacrificed to this end.

"Intuitiveness" is the less scientifically recognized inclination of this classification in spite of the fact that some individuals are sometimes strongly motivated by intuition. Our vocabulary does not offer specific words to characterize the low states of development of this strange personality "dimension." One is with increased intuition *"spiritless,"* "semiintuitive," "intuitive," or "psychic." A person is considered psychic when his intuitiveness reaches quite the pathological level. Such a person is often distressed by his sensitivity to impulses whose origin transcends the field of sensorial perception and of mental concepts. The knowledge acquired by intuition can sometimes be translated into understandable information through the play of the imagination.

Another strange aspect of this inclination is that an unintuitive person can be so thoroughly "blinded" to psychic impulses that he is incapable of recognizing even the slim possibility of intuitive manifestations.

H. COMPOUND ATTITUDES

As stated above, the currently identified character traits correspond to compound attitudes, as shown in Table 2. A misanthrope, for example, who hates mankind is, according to this classification, a person who is simultaneously "distant," "unconcerned," and resentful, while a haughty person, disdainfully proud, is together "discriminating," "unconcerned," and "vain."

TABLE 2
EXAMPLES OF COMPOUND ATTITUDES

Character trait	Elementary components				
Gourmet	Gourmand	+	"Refined"		
Bon vivant	Gourmand	+	Sociable		
Epicurian	Gourmand	+	Sensual	+	"Refined"
Jealous	Possessive	+	"Unconcerned"		
Envious	Acquisitive	+	"Unconcerned"		
Grudging	Acquisitive	+	"Unconcerned"	+	Resentful
Haughty	"Discriminating"	+	"Unconcerned"	+	"Vain"
Misanthrope	"Distant"	+	"Unconcerned"	+	Resentful
Cynical	"Discriminating"	+	Uncooperative	+	Resentful

I. INDIVIDUAL RESISTANT PROFILE

The well known personality of General de Gaulle will be taken as an example to show how a psychological profile can be established in the light of the above classification.

Charles de Gaulle has often shown fearlessness and haughtiness in his career. The latter attitude suggests, according to Table 2, that he is "discriminating," "unconcerned," and "vain." He possesses a highly conceptive mind, having predicted the importance of a fully mechanized army before

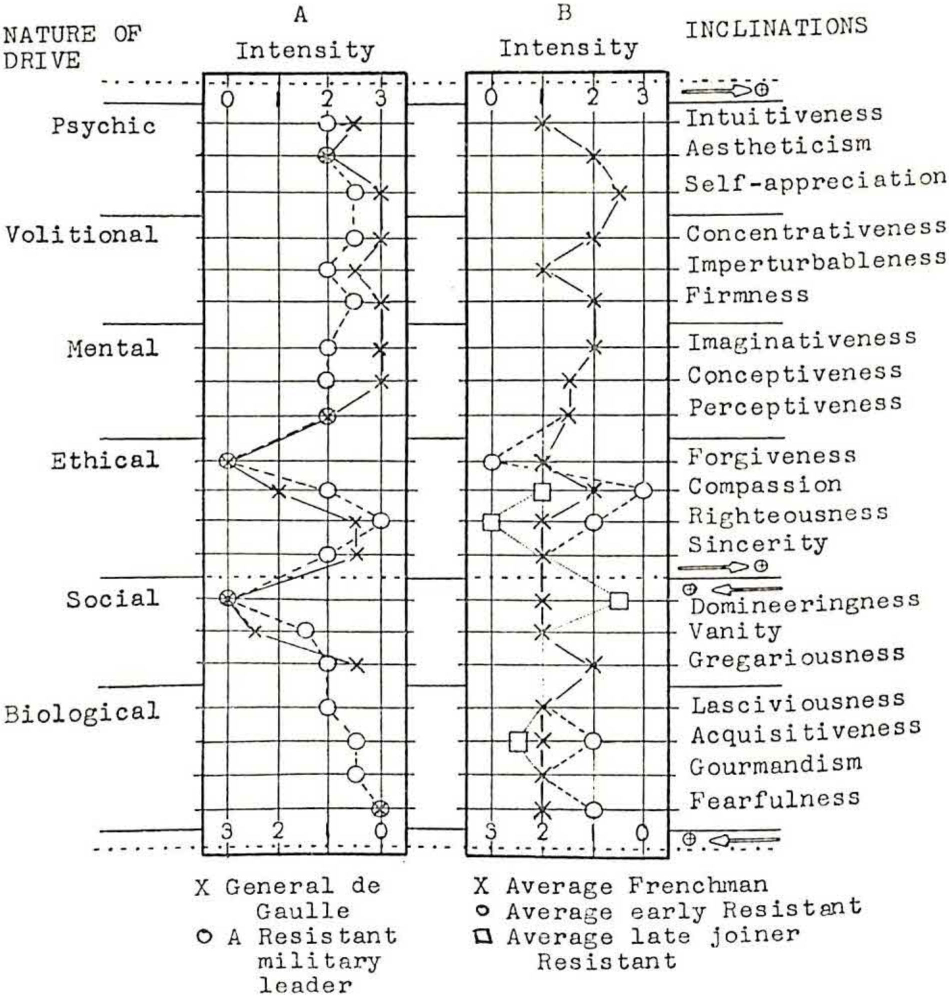


FIGURE 2
PERSONALITY PROFILES

World War II and conceived how to create a political Europe extending from the Atlantic to the Urals under the nonapparent ruling of France. De Gaulle is also well known for his stubbornness and his ability to bring a program to completion. These aptitudes give him a high score in concentrativeness and firmness. He often did the right thing at the right time, a symptom of intuitiveness. All these character traits are represented by crosses in the personality profile shown by Diagram A of Figure 2.

The personality profile of one of the local military leaders of the French resistance is shown by the circles in the same diagram. The similarity in the psychological profile of these two men is striking. Their outstanding accomplishments result from the high score in their mental and volitional abilities. Their success as leaders results from their strong domineering inclination, their absence of fear, and their strong sense of righteousness.

J. Composite Resistant Profile

The personality of an average resistant was not essentially different from the personality of an average Frenchman. The profile of the latter is given by the crosses of Diagram B, Figure 2. As shown by this diagram, the average Frenchman is cautious, "gourmand," possessive, and sensual; "discriminating," "vain," and uncooperative; insincere, partial, humane, and resentful; "aware," reasonable, and imaginative; perseverant, emotional, and attentive; "sure of himself," "refined," and semiintuitive.

The first "resistance" fighters, those who started the movement soon after the surrender of France, were usually strongly motivated by a sense of justice and compassion. They were deeply convinced that Hitler's domination views were incompatible with their way of life, which involved the respect of individual liberty and of humane dignity. The psychological profile of these fighters is given by the circles of Diagram B. This profile differs from the profile of the average Frenchman by stronger righteousness, compassion, and fearlessness, and by weaker forgiveness.

Those who joined the resistance afterward were in majority people who feared the labor draft to Germany. These "rebels" were moved more by social and biological than by ethical drives. They preferred the risks of the "maquis" life to the hardship of slave labor under hostile environment and constant fear of bombardment. Most of them were revengeful, brave, and, before all, nonsubmissive.

Those who joined when the end of the war was in sight were mostly opportunists. Many of them had compromised with the occupant. They were now willing to take the risks of the "maquis" to rehabilitate themselves. They were distinctly submissive, partial, and "unconcerned."

K. CONCLUSION

Two decades have passed since the largest portion of this study was achieved. Improvements in the classification of elemental attitudes were made by extending the authors' observations to individuals living under more normal conditions, and by discussing their point of view with numerous people, including professional psychologists.

The findings resulting from this study have acquired a broader significance than that of discovering the psychological motivation of resistance fighters. In fact, this study leads to a new way of classifying basic behavior patterns and gives a matrix which allows the accurate description of the personality profile of any individual. From a linguistic point of view, the essay ascribes an unequivocal use to many terms of the unprecise character vocabulary and establishes meaningful relations between some of these terms. Finally, from a philosophical and metaphysical point of view, it rehabilitates ethical and psychical values in the field of psychology, which is a domain of knowledge wherein considerations stemming partially from subjective feelings cannot be ignored.

REFERENCES

1. DELMAS, A., & BOLL, M. La Personalité Humaine, Son Analyse. Paris: Flamarion, 1922.
2. EYSENCK, H. J. The Scientific Study of Personality. London: Routledge & Kegan Paul, 1952.
3. MURRAY, H. A. Exploratory in Personality. New York: Oxford Press, 1933. Pp. 144-150.

Green Acres
Bryn Mawr, Pennsylvania 19010

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 281-285.

INTERPERSONAL ATTRACTION AS A FUNCTION OF FREEDOM OF CHOICE* [1]

The University of Texas at Austin

S. THOMAS FRIEDMAN AND DAVID SHERRILL

A. INTRODUCTION

Various factors enhancing or inhibiting attraction between two persons have been systematically reported. Byrne and Nelson (3), Aronson and Linder (1), Keisler (5), Newcomb (6), and Rokeach (7, 8), among others, have examined dimensions of similarity and dissimilarity that influence the interpersonal attraction process.

The problem can be cast into another frame of reference to ask the question: What are the differential response patterns under different degrees of freedom of choice? In this study, the role of roommate was selected and the female *S*s were asked to indicate the kinds of people they would "choose" as roommates as well as the degree to which they would "accept" certain characteristics in their roommates.

B. METHOD

Subjects were 1032 female students at The University of Texas (Austin), almost the entire female population of five University-sponsored dorms (the only missing girls were those ill or away from school for personal reasons). The students anonymously completed a 20-item series asking degree of "choosing" and degree of "accepting" as roommates girls with different characteristics. Using a five-point scale ranging from strongly agree to strongly disagree, the *S*s indicated the degree of acceptance of each characteristic in "choosing" or "accepting" a roommate.

C. RESULTS

Table 1 compresses the five-point rating scale into a trichotomy; agreement or acceptance, undecided, disagree or rejection. It is obvious, from inspection of Table 1, that a significantly larger number of coeds say they would accept

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 11, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This study was supported in part from funds provided by the Harrell Ingham Study of Socio-Religious Attitudes of University Students, sponsored by the Methodist Campus Ministry through the Wesley Foundation, The University of Texas, Austin.

as roommates girls of differing backgrounds than would voluntarily choose such girls. This held for every differing group—religious, political, point of origin, ethnic, and racial (*t*-test for correlated means).

TABLE 1
PROPORTIONS OF FEMALE STUDENTS CHOOSING AND ACCEPTING ROOMMATES OF DIFFERING BACKGROUNDS
($N = 1032$)

Item	Proportion expressed in percentage 1 Agree[a]	 2 Undecided	 3 Disagree
1. Choose similar social and economic background	71	13	6
2. Accept similar social and economic background	96	3	1
3. Choose different nationality	18	38	44
4. Accept different nationality	52	27	22
5. Choose Protestant faith	71	22	8
6. Accept Protestant faith	97	3	0
7. Choose Catholic faith	57	30	12
8. Accept Catholic faith	94	4	2
9. Choose Jewish faith	36	38	26
10. Accept Jewish faith	74	16	10
11. Choose another faith	54	34	12
12. Accept another faith	84	13	3
13. Choose different political thinking	50	38	12
14. Accept different political thinking	87	12	3
15. Choose Oriental student	20	43	37
16. Accept Oriental student	59	24	17
17. Choose Latin American student	18	37	45
18. Accept Latin American student	49	25	26
19. Choose Negro student	8	22	70
20. Accept Negro student	26	20	54

[a] All differences significant at .01 level; *t*-test for correlated means.

These findings lend confirmation, in the area of interpersonal relations, to the general social-psychological hypothesis that individuals will accept a larger number of options in interpersonal relations than they would voluntarily choose.

Choosing "like" roommates leads to a hypothesized minimum of adjustment and emotional output. Accepting "different" roommates implies an element of curiosity, novelty, and learning that the individual is willing to undertake, but for which he will not necessarily volunteer.

In the attempt to obtain patterns of similarity and differences among subgroups of the total population ($N = 1032$), the 20-item scale was subjected to factor analytic techniques. After posting the responses for each S to IBM cards, the intercorrelations for each item with every other item were obtained. The resulting 20×20 correlation matrix was subjected to factor extraction by the principal axis method.

Those extracted factors having eigenvalues of 1.00 or higher were subsequently rotated to orthogonality by means of Kaiser's (4) normalized varimax method.

Table 2 presents those items loading .30 or higher on each factor, the

TABLE 2
PRINCIPAL AXIS (VARIMAX ROTATED) FACTORS OBTAINED FOR 20-ITEM ATTITUDE QUESTIONNAIRE
($N = 1032$)

Loading	Item
Factor I Ethnic Openness 27.06% variance	
.84	Choose Latin American
.81	Choose Negro
.81	Accept Latin American
.81	Accept Negro
.78	Choose Oriental
.75	Choose Different Nationality
.74	Accept Different Nationality
.74	Accept Oriental
.43	Choose Jewish
.39	Accept Jewish
Factor II Seeking Religious Openness 13.27% variance	
.75	Accept another faith
.74	Choose different political thought
.70	Choose Catholic
.56	Choose Jewish
.47	Choose Protestant
.37	Choose Oriental
.31	Choose Latin American
Factor III Acceptance of Differences 14.22% variance	
.76	Accept another faith
.73	Accept Jewish faith
.71	Accept different political thought
.42	Accept Oriental
.33	Choose Jewish
.32	Accept Protestant
.32	Accept Latin American
.30	Accept different nationality
Factor IV Social Class Factor 7.44% variance	
.82	Choose same SES
.80	Accept same SES
Factor V Christianity Factor 7.97% variance	
.81	Accept Protestant
.65	Choose Protestant
.46	Accept Catholic
.30	Choose Catholic

percentage of the variance accounted for by the factor, and a tentative name for the factor. At the time the *S*s completed the 20-item scale they indicated their year in school, college in the university, religious preference, and sorority affiliation or nonaffiliation.

D. Discussion

Factor scores were computed for each *S*, and single classification analyses of variance were computed for each grouping. Seniors are significantly more "ethnically open," more receptive to religious openness, and place less emphasis on social class factors than lower classmen or junior students. A possible selective factor, of who stays in college, could also be suggested as the reason for these results as well as the impact of the college experience.

When factor scores are compared by religious affiliation of the respondents, significant differences (.001) appear for each of the five factors. Members of minority religions, Catholic and Jewish students, are more accepting of religious and racial differences (Factors I, II and III) than are other students. The Jewish students appear to be more sensitive to social class items and, as expected, reject Factor V, the Christianity factor. Jewish students are more accepting of differences than either Protestant or Catholic students.

When grouped by college of enrollment, two of the five factors, Ethnic Openness and Seeking Religious Openness, yield significant differences. Business school majors are the most conservative, Education majors next, and Arts and Sciences and Fine Arts students most "open" and "seeking."

Lastly, the *S*s were grouped by Greek letter affiliation. The two groups differ significantly on Factors I and IV. Sorority members are more sensitive to social class background (Factor IV), a tribute to the sorority selection process or training procedures or both; and on Factor I, Ethnic Openness, independent girls score significantly higher than sorority members.

To sum up, the evidence suggests that female students state that they will accept a wider degree of variation in roommates than they would voluntarily choose. Factor analysis of the 20-item scale identifies several dimensions across which significant differences among students as a function of length of time in the university, college enrolled in, religious preference, and Greek letter affiliation are found.

E. Summary

What differences in response patterns of interpersonal attraction are indicated when female students are given the option of voluntarily choosing a roommate or are not given this option and must accept the roommate? The

subjects ($N = 1032$) were coeds housed in five university-supervised dorms at The University of Texas (Austin). The data clearly demonstrate that these students will accept, in significantly greater numbers, roommates of diverse and different backgrounds than they would voluntarily choose. This applies to roommates of differing origin, socioeconomic status, religion, and race.

Analysis of subgroups of the total population based on factor analytic techniques indicates that length of time at The University of Texas, religion of the student, college enrolled in, and Greek letter affiliation clearly differentiate "Ethnically Open" students and sensitivity to factors of "Seeking Religious Openness," "Acceptance of Differences," "Social Class," and "Christianity."

REFERENCES

1. ARONSON, E., & LINDER, O. Gain and loss of esteem as determinants of interpersonal attractiveness. *J. Exper. Child Psychol.,* 1965, **1**, 156-171.
2. BYRNE, D. Interpersonal attraction and attitude similarity. *J. Abn. & Soc. Psychol.,* 1961, **62**, 713-715.
3. BYRNE, D., & NELSON, D. Attraction as a linear function of proportion of positive reinforcements. *J. Personal. & Soc. Psychol.,* 1965, **1**, 659-663.
4. KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika,* 1958, **23**, 187-200.
5. KEISLER, E. R. Experimental development of "like" and "dislike" of others among adolescent girls. *Child Devel.,* 1961, **32**, 59-66.
6. NEWCOMB, T. M. The production of interpersonal attraction. *Amer. Psychol.,* 1956, **11**, 575-586.
7. ROKEACH, M. The Open and Closed Mind. New York: Basic Books, 1960.
8. ———. Belief versus race as determinants of social distance: Comments of Triadris' paper. *J. Abn. & Soc. Psychol.,* 1961, **62**, 187-188.

Department of Educational Psychology
The University of Texas
Austin, Texas 78712

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 287-289.

USE OF THE EDWARDS PERSONAL PREFERENCE SCHEDULE WITH ALCOHOLICS IN A VOCATIONAL REHABILITATION PROGRAM*

Texas Research Institute of Mental Sciences, Houston, Texas

ALLEN GOSS,[1] TOM MOROSKO, AND ROBERT SHELDON

A. INTRODUCTION

Of the reported studies in which the Edwards Personal Preference Schedule, EPPS (1), has been used, only one has been with alcoholic patients (2). Results from other recent studies indicate that factors, such as age (3), population (4), and diagnoses (5), are important variables to consider in analyzing EPPS data. This study compares the 15 EPPS score areas between Edwards General Adult Sample and alcoholic subjects accepted into a vocational rehabilitation program, and also compares the scores for the patients in the program who gained employment with those patients who did not gain employment.

B. METHOD

*S*s were 18 male neuropsychiatric patients accepted into the Vocational Rehabilitation Program at the VA Hospital, Houston. The mean age, *IQ*, and education for the group were, respectively, 44, 108.8, and 11.1. These are similar to the values reported by Fitzgerald, Pasewack, and Tanner (2) in their EPPS alcoholic study.

The *S*s were administered the EPPS at the time of acceptance into the vocational program. All *S*s had taken part in other treatment programs at the VA Hospital prior to being transferred to the Vocational Rehabilitation Program and had, therefore, been dry a considerable time prior to acceptance. Data were also collected on a number of demographic variables. The criterion of employment in this study was related to the discharge status of the patient. If the patient was discharged from the ward with suitable employment, he was considered a "success." If he was transferred or discharged from the ward in any other category, he was considered a "failure."

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 12, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] The author was formerly at the VA Hospital, Houston, Texas. Appreciation is extended to Kenton Pate, who is in charge of the Vocational Rehabilitation Program.

C. RESULTS AND DISCUSSION

The analysis of 10 demographic variables indicated nonsignificant differences between the success and failure groups. Table 1 presents the percentile scores for both the success and failure groups as well as for the combined alcoholic population for the 15 EPPS need variables. A rank ordering of the need scores is also presented as well as the results from the analysis of the success and failure group scores for the 15 variables.

TABLE 1
EDWARDS PERSONAL PREFERENCE SCHEDULE PERCENTILE SCORES AND RANKS FOR AN ALCOHOLIC POPULATION IN A REHABILITATION PROGRAM

EPPS variable	Total $N = 18$	Rank	Success $N = 10$	Rank	Failure $N = 8$	Rank	S-F P
ACH	71.2[a]	1	70.30	1	72.37	2.5	.90
DEF	56.3	5	44.40	12	71.25	4	.05
ORD	47.4	10	39.50	14	57.37	6	.27
EXH	46.4	13	45.90	11	47.00	8	.94
AUT	49.6	7	53.10	7	45.12	9	.56
OFF	47.0	12	53.80	6	38.50	12	.29
INT	66.6	2	60.60	3	74.12	1	.32
SUC	49.9	6	62.30	2	34.50	14	.04
DOM	43.2	14	43.40	13	43.00	10	.99
ABA	42.7	15	52.50	9	30.37	15	.20
NUR	47.1	11	56.50	5	35.25	13	.17
CHG	58.6	4	52.70	8	65.87	5	.34
END	48.4	8	46.40	10	51.00	7	.75
HET	65.2	3	59.40	4	72.37	2.5	.27
AGG	47.6	9	51.50	9	42.62	11	.48

[a] Percentile scores from Edwards General Adult Group.

The results indicate that score areas which vary the most from the general adult norms are not necessarily the areas of greatest significance for the employment criterion. For example, the achievement score for the alcoholic population is at the 71st percentile and yet it does not distinguish between the success or failure group at all ($F = .02$, $p = .90$). The two score areas which differentiate the employed from the unemployed at accepted levels of significance are deference ($S = 44.40$, $F = 71.25$; $F = 4.55$, $p = .05$) and succorance ($S = 62.30$; $F = 34.50$; $F = 5.00$, $p = .04$). While not separating the groups at conventionally accepted statistical levels, two other score areas may prove helpful in separating the populations; these areas are nurturance and abasement.

A comparison of the present results with the results of Fitzgerald, Pasewack, and Tanner (2) indicates little similarity. There were two important differences between the designs of the studies. The major concern of the Fitzgerald *et al.* (2) study was an analysis of the mean difference between alcoholic patients and the general adult norms. The present analysis would seem to indicate that large mean differences between alcoholic populations and the general adult norms are not necessarily the difference of greatest practical importance, but rather the mean differences which exist between pragmatically important criterion groups. Since mean differences between the treatment completion and noncompletion groups or the readmission and nonreadmission groups were not presented by Fitzgerald *et al.* (2), criterion comparisons were impossible.

D. SUMMARY

With the EPPS, comparisons between the Edwards General Adult Sample and alcoholic patients in a Vocational Rehabilitation Program were shown to be relatively unimportant with respect to their vocational outcome. A comparison of EPPS scores for patients who gained employment with those patients who did not was shown to provide useful information. Since the areas of greatest difference between alcoholics and normals were not necessarily the areas of importance in separating the alcoholic population on a pragmatic criterion, employment, the results were regarded as support for a continuing need to investigate practical behavioral outcomes.

REFERENCES

1. EDWARDS, A. L. Edwards Personal Preference Schedule Manual. New York: Psychol. Corp., 1959.
2. FITZGERALD, B. J., PASEWACK, R. A., & TANNER, C. E. Use of the Edwards Personal Preference Schedule with hospitalized alcoholics. *J. Clin. Psychol.*, 1967, **23**, 194-195.
3. GAURON, E. F. Changes in Edwards Personal Preference Schedule needs with age and psychiatric status. *J. Clin. Psychol.*, 1965, **21**, 194-196.
4. GOSS, A. Predicting work success for patients on an industrial rehabilitation ward in a neuropsychiatric setting. Doctoral dissertation, University of Texas, Houston, Texas, 1966. University Microfilms, Ann Arbor, Michigan, 1966, No. 66-14381.
5. NOWICKI, S. Use of the Edwards Personal Preference Schedule in a psychiatric population. *J. Clin. Psychol.*, 1967, **23**, 361-362.

Texas Medical Center
1300 Moursund Avenue
Houston, Texas 77025

Published as a separate and in *The Journal of Psychology*, 1968, 68, 291-298.

THE CORRELATION OF TWO SELF-DISCLOSURE INVENTORIES WITH ACTUAL SELF-DISCLOSURE: A VALIDITY STUDY*

Department of Psychology, Brigham Young University

DARHL M. PEDERSEN AND VINCENT J. BREGLIO

A. INTRODUCTION

Self-disclosure, as defined by Jourard and Lasakow (5), is the process of making the self known to other persons. Jourard has designed several different inventories to measure self-disclosure (3, 4, 5). In each a separate measure of disclosure was obtained for each of several different target persons (mother, father, best friend, spouse, etc.). Also the sum of the disclosure scores for the different target persons was taken to obtain an overall self-disclosure score. In a recent study (4), self-disclosure scores were used to predict overall achievement during a four-year nursing program. The results indicated that the degree to which the subjects said they had engaged in self-revealing communication to significant others in their lives was highly related to obtaining good grades in nursing college. These findings have provided some evidence for the predictive validity of self-disclosure questionnaires and have suggested that self-disclosure may be an important nonintellective factor predictive of success in such professions as counseling, psychotherapy, teaching, etc.

Other studies designed to demonstrate predictive, concurrent, or construct validity for the self-disclosure scales have not been very successful. Lubin and Harrison (6), using participants in a group-process conference, found that trainer ratings of self-disclosing behavior assessed after 20 hours of group interaction were not significantly correlated with total score on the Jourard Self-Disclosure Inventory. Himelstein and Kimbrough (1) obtained nonsignificant correlations between total score on the original self-disclosure inventory of Jourard and the amount of information disclosed by graduate students making personal introductions in a classroom situation. One attempt to validate a version of the Jourard Self-Disclosure Inventory involved the peer-nomination technique (2). The inventory was correlated with the judgment by peers of the extent to which the individual was likely to confide in others or be

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 15, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

confided in by others. The results indicated no significant relationships of the Self-Disclosure Inventory with peer nomination scores for either the likelihood of confiding or being confided in.

This study represents an effort to establish further the validity of two self-disclosure scales of Jourard. The inventories which measure the extent to which the subject states that he has already discussed a number of items with four target persons will be correlated with the actual disclosure of the subject about five topics.

B. METHOD

1. *Subjects*

The subjects for this study were 52 students (26 males and 26 females) enrolled in two introductory psychology courses at Brigham Young University.

2. *Measuring Instruments*

a. Jourard Self-Disclosure Inventory (SD-60). The SD-60 was used by Jourard in early studies (5). It contains 10 statements in each of six subject areas: attitudes and opinions, tastes and interests, work (or studies), money, personality, and body. The subjects in this study were asked to indicate the extent to which they had discussed the 10 statements under each topic with each of four target persons—mother, father, best male friend, and best female friend. The subjects used the following scale:

1. Have told the other person nothing about this aspect of me.
2. Have talked in general terms about this item. The other person has only a general idea about this aspect of me.
3. Have talked in full and complete detail about this item to the other person. He knows me fully in this respect and could describe me accurately.

The complete inventory may be found in the article by Jourard and Lasakow (5).

b. Shortened Version of the Jourard Self-Disclosure Inventory (SD-25). The SD-25 is another self-disclosure questionnaire, containing 25 questions, which was used subsequently by Jourard (4). The statements tend to cover the same six topic areas as those included in the SD-60, but they are worded differently and are not explicitly divided into topic areas. The task of the subjects was the same as for the SD-60. They were instructed to indicate the extent to which they had discussed each statement with each of the same four target persons—mother, father, best male friend, and best female friend. The following scale, which differs from that used in the SD-60, was used:

1. Have not disclosed this information to that person or have given an incomplete version of the information.
2. Have disclosed the information to the other person.

The complete inventory may be found in the article by Jourard (4).

c. Self-Disclosure Questionnaire (SDQ). The SDQ consisted of five questions, each presented at the top of a separate sheet of paper. Each question was constructed to cover one of the areas covered by the SD-60, although two of the questions covered only part of an area. Rather than attempt to include both tastes and interests, the question was worded to elicit disclosure of interests only. Also the question on body was restricted to cover the subtopics of health and physical appearance. A question on attitudes and opinions was not included because of the large overlap of the subject matter with interests. The areas covered and the questions representing them are as follows:

1. *Interests.* What are your main interests in life? Describe those activities and experiences which you enjoy.
2. *Personality.* Describe your personality. Consider both your strengths and weaknesses.
3. *Studies.* Describe yourself as a student. Include in your description an evaluation of your abilities, current motivation, and future educational and/or occupational plans.
4. *Body.* Describe your present state of health and physical appearance. What do you like and/or dislike about your physical characteristics?
5. *Money.* What is your current financial condition? Include a description of both your financial assets and liabilities.

3. *Procedure*

The SD-60 and SD-25 were administered during a one-hour class period to each subject. Eleven weeks later, the SDQ was administered to each subject. To minimize the association by the subject between the two testing situations, different experimenters were used.

Responses to the SD-60 and SD-25 were recorded by each subject on IBM answer sheets. The sheets were fed through an IBM 1230 scanner which automatically scored them and punched summary scores into IBM cards. Five scores on each of the two instruments were computed for each subject. They were disclosure to mother (M), disclosure to father (F), disclosure to best male friend (BMF), disclosure to best female friend (BFF), and total disclosure. Seven scores of actual self-disclosure were obtained from the SDQ for each subject. A depth of disclosure score was derived for each of the five questions by the following procedure. First, depth of disclosure to each of the five questions was rated by two independent raters using the following three-point scale:

1. Discloses little or nothing of himself.
2. Discloses only on a general level; discloses facts but little or no disclosure of a personal nature.
3. Discloses attitudes and information of a personal nature.

Second, since the rater reliability coefficient of agreement for the depth of disclosure ratings was .83, the independent ratings were summated to yield a single depth of disclosure score for each question. A total depth of disclosure score was obtained by summating the separate depth of disclosure scores for each of the five questions. A total amount of disclosure score was obtained by counting the number of words used to answer the five questions.

C. Results and Discussion

The correlations between the 10 SD-60 and SD-25 self-disclosure scores and the seven self-disclosure scores from the SDQ are presented in Table 1. A glance at the table reveals some very interesting trends. The largest correlations are in the row involving self-disclosure about studies on the SDQ. The two correlations between actual self-disclosure about studies and claimed disclosure to mother and father on the SD-60 are moderately high—.47 and .49, respectively. The correlations between disclosure about studies and claimed disclosure to best male and female friends on the SD-60 are also significant although not as large—.28 and .29, respectively. With these correlations it is not surprising that the correlation of studies with total disclosure on the SD-60 is significant, since the total disclosure score is simply the sum of the separate target person disclosure scores. That is, correlations involving the total disclosure score on the two self-disclosure inventories are not independent of the correlations involving the separate target person disclosure scores.

The same pattern of correlations is evidenced between studies and the SD-25 disclosure scores. However, they are smaller in magnitude. There are two factors which may have attenuated those correlations. First, since the SD-25 has less than half the items of the SD-60, it probably has a lower reliability. Second, the SD-25 and the SD-60 may be measuring somewhat different attributes. Both factors are probably involved, since not only do the correlations involving the SD-25 tend to be lower, but also there are somewhat different patterns of correlations involving the two inventories.

Even though the pattern of correlations between studies and the SD-25 disclosure measures is the same as for the SD-60, the correlations between studies

TABLE 1

CORRELATIONS OF THE SD-60 AND SD-25 MEASURES WITH THE SDQ MEASURES

	SD-60					SD-25				
SDQ	M	F	BMF	BFF	Total	M	F	BMF	BFF	Total
Interests	.12	.15	.16	.20	.21	.16	.19	.05	.19	.20
Personality	.21	.13	.26	.16	.26	.20	.07	.17	.07	.17
Studies	.47*	.49*	.28*	.29*	.50*	.34*	.31*	.18	.06	.28*
Body	.20	.10	.06	.05	.13	.32*	.17	.03	.20	.23
Money	.20	.21	—.17	—.03	.05	.19	.10	—.16	—.16	—.02
Total depth	.34*	.31*	.17	.19	.32*	.35*	.24	.08	.11	.25
Total amount	.34*	.28*	.19	.10	.29*	.31*	.19	.11	.06	.21

Note: $N = 52$.

* $p < .05$.

and disclosure to BMF and BFF are not significant. As the correlations are lowered on the SD-25 as compared to the SD-60, these two drop below significance. The conclusion from the consistent correlations between actual disclosure about studies and the claimed self-disclosure measures is that people who state that they had disclosed relatively more about themselves to various target persons tend actually to disclose more about themselves in the testing situation only on the topic of studies. Incidentally, the significant correlation between actual disclosure about body and disclosure to the mother on the SD-25 cannot be given very much credence because the corresponding correlation for the SD-60 is not significant. Furthermore, it would be difficult to develop a rationale to explain that relationship.

The question may be raised as to why studies was the only topic of actual self-disclosure related to disclosure on the SD-60 and the SD-25. One explanation may be that the other topics are more personal, and the subject may be reluctant to disclose information about them to an experimenter whom he did not know. Therefore, disclosure about studies may be the only topic that a person claims he has discussed with the various target persons that he would be willing actually to disclose to the experimenter. A better test of the relationship between claimed and actual self-disclosure would be to obtain measures of both directed toward the same target persons. However, it is still possible that actual and claimed self-disclosure are correlated only for the area of studies even for all four of the target persons considered. Furthermore, it was the intention of the present research to determine the generalizability of the characteristic of self-disclosure as measured by the Jourard inventories to other situations involving self-disclosure.

Other correlations in the matrix which require attention are the total depth and the total amount of disclosure correlations with the claimed self-disclosure measures. One of the most noticeable aspects of these correlations is that the patterns and magnitudes of the correlations for total depth and total amount are almost identical. This is primarily due to the high positive correlation that was found to exist between total depth and total amount (.84). Prior to the experiment it was anticipated that there would be a moderately high correlation between the rating of the depth of the person's disclosure and the total number of words he would use in the disclosure; however, it was not expected that it would be of this magnitude. This finding has some valuable implications for research. Since the intercorrelation is of the magnitude of an acceptable reliability coefficient, amount of disclosure may be substituted for depth of disclosure. Amount of disclosure is a much more immediate, objective, and reliable measure, and it is much more useful in many conceivable research applications. Using this measure, reinforcement schedules can be programmed to investigate the influence of various forms of reward and punishment on an individual's self-disclosure in both individual and group situations. The effects of reinforcement on patterns of interrelations among the self-disclosure of two or more persons can be examined.

Another salient feature of the correlations between total depth and total amount of disclosure with the SD-60 and the SD-25 measures is that the patterns of correlations tend to parallel the pattern between actual disclosure about studies and the SD measures. The correlations between total depth on the SDQ with the self-disclosure inventory measures are primarily a result of the fact that total depth is derived by summing the depth ratings for each of the separate topics so that total depth is not independent of studies. Therefore, the correlations involving total depth reflect primarily the relationships between studies and the other variables. The total depth correlations tend to be smaller because, when the ratings of depth of disclosure to the other topics are added to the rating of depth of disclosure to the topic of studies to yield the total depth of disclosure rating, they tend to dilute the contribution of depth of disclosure rating of the topic of studies. The correlations involving total depth corresponding to the smaller significant correlations found for studies tend to be reduced below the level required for significance. Nonetheless, it is true that both total depth and total amount of disclosure as measured by the SDQ are significantly correlated with total self-disclosure as measured by the SD-60. People who claim on the SD-60 to have disclosed more to target persons tend actually to disclose more on a questionnaire which requests them to disclose information about themselves. This finding provides some construct validity for the SD-60. Unfortunately, perhaps because of its relative unreliability, similar evidence for the construct validity of the SD-25 was absent. It should be noted that total depth and total amount also correlated significantly with disclosure to mother and father on the SD-60 and disclosure to mother on the SD-25. Again it should be emphasized that these latter correlations are primarily derived from the actual disclosure of the studies component which they possess.

D. SUMMARY

Two self-disclosure measures developed by Jourard (SD-60 and SD-25) and a Self-Disclosure Questionnaire (SDQ) were administered to 52 *S*s. Each of the Jourard measures yielded four scores which indicated the extent to which the *S*s *claimed* they had disclosed information about themselves to four target persons—mother (M), father (F), best male friend (BMF), best

female friend (BFF)—and a total disclosure score (sum of the four target person disclosure scores). The SDQ yielded separate scores of *actual* depth of self-disclosure in five topic areas (interest, personality, studies, body, and money), a total depth of disclosure score (sum of the separate depth scores), and a total amount of disclosure score (count of the words used to answer the five questions).

Correlations between the 10 Jourard scores and the seven SDQ scores indicated that (*a*) total depth of disclosure was highly correlated with total amount of disclosure (.84); (*b*) both total depth and total amount of disclosure were correlated with M, F, and total disclosure on the SD-60 and M on the SD-25; (*c*) the only area of actual disclosure consistently related to claimed disclosure was that of studies; and (*d*) SD-60 and SD-25 measure somewhat different aspects of self-disclosure.

REFERENCES

1. HIMELSTEIN, P., & KIMBROUGH, W. K. A study of self-disclosure in the classroom. *J. of Psychol.*, 1963, **55**, 437-440.
2. HIMELSTEIN, P., & LUBIN, B. Attempted validation of the self-disclosure inventory by the peer-nomination technique. *J. of Psychol.*, 1965, **61**, 13-16.
3. JOURARD, S. M. Self-disclosure and other-cathexis. *J. Abn. & Soc. Psychol.*, 1959, **59**, 428-431.
4. ———. Self-disclosure scores and grades in nursing college. *J. Appl. Psychol.*, 1961, **45**(4), 244-247.
5. JOURARD, S. M., & LASAKOW, P. Some factors in self-disclosure. *J. Abn. & Soc. Psychol.*, 1958, **56**, 91-98.
6. LUBIN, B., & HARRISON, R. L. Predicting small group behavior with the self-disclosure inventory. *Psychol. Rep.*, 1964, **15**, 77-78.

Department of Psychology
Brigham Young University
Provo, Utah 84601

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 299-303.

INSTITUTIONALIZED RETARDATES' REACTIONS TO AUTHORITY FIGURES: GOD AND MONSTER DRAWINGS*

Willowbrook State School, Staten Island

GABRIEL PUSTEL, MANNY STERNLICHT, AND LOUIS SIEGEL

A. INTRODUCTION

Both Freud and Reik's view of religion is that God and the Devil (and Monsters, by implication) are psychologically a splitting of "good" and "bad" facets of a magnified father image (1, 5). The Oedipal theory posits, in the Judaeo-Christian tradition, a greater fear of and hostility toward God and Devil (Monster) by males than by females. This attitude, to a great extent, represents a displacement of original attitudes toward parents or parental surrogates. With institutionalized individuals, the institution itself may stand as an authority figure. If the life experiences of such institutionalized patients have been largely of a negative type, we would expect them to project a negative attitude onto God and Monster drawings.

We also would expect those institutionalized patients with weak ego development, particularly adolescents, to evidence greater signs of pre-Oedipal, oral-sadistic cimmerian fantasies, and that, in light of Klein's findings (2, 3), such fantasies would be readily projected onto Monster drawings. Adults would be expected to demonstrate less indications of cimmerianism, because of greater control of "internal good and bad objects" via processes of introjection, incorporation, and identification. Further, because of their still viable oral-sadistic tendencies, adolescents might be expected to exhibit more animalistic reactions to Monsters, with adults highlighting more humanoid-type Monsters.

Institutionalized adolescents also might reveal larger drawings of Monsters than of God, especially in comparison with adults, due to a more fluid, alternating "good" and "bad" introjection of protective and persecutory object representation. Fear of retaliation and a less well-integrated ego may make adolescent *S*s more prone to being overwhelmed by the complexity of a frustration-aggression learning dilemma. They are in a motivational state involving

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 19, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

greater expectations of gratification from the potential tension-reducing authority figures of the institution. On the other hand, adults, with their greater experience in terms of negative *vs.* positive motivational conflict and subsequent developmental identifications, adapt their levels of aspiration to a lower level of reinforcement, via the utilization of perceptual defenses. That is, adaptive behavior involves a certain blunting of affect, with a relative lack of hope in bringing back "good" feelings within the social context of institutional living.

Thus, the symbolic equivalence of God and Monster drawings (the splitting of the father figure into a "good" and "bad" affect dimension) among adolescent and adult institutionalized mental retardates should inferentially reveal levels of adaptation and aspiration to a parental surrogate (e.g., the institution) as a progressive learning experience in which, for the adult *S*s, cue-producing responses lose a great deal of their cathexis and are replaced by the perceptual defense of developmental indifference (as a response to the accumulations of the ongoing frustrations of institutional life).

Lederer (4), in a fascinating study, feels that psychoanalysis currently overemphasizes the matriarchal aspects of superego development, and he tries to correct this by stressing the patriarchal influences. He describes occidental dinosaur monsters as having "fat rumps, short legs," as having "a lizard's head with carnivorous teeth," as being "bovine," and Lederer interprets these characteristics as being masculine. While "carnivorous teeth" may very well be symbolic of masculinity, "fat rumps" and "bovine" would seem to be more suggestive of femininity.

We believe that the dinosaur may be viewed as a feminine-masculine animal —that is, a transitional animal—insofar as symbolic ego identification is concerned. In general, as the ego tends to become more fully integrated, there is a shift from feminine to feminine-masculine to masculine identifications. However, depending upon the degree of ego identifications, this does not necessarily exclude the coexistence of feminine-masculine identifications.

B. METHOD

The *S*s of this experiment consisted of a total of 60 institutionalized (at Willowbrook State School) *S*s, dichotomized into an adolescent group (CAs 12-20, with a mean CA of 16.5) and an adult group (CAs 21-50, with a mean CA of 29.8). Total *IQ*s ranged from 52 to 74, with mean *IQ*s of 64.2 for the adolescent *S*s and 62.6 for the adults. Each age group was further subdivided into two equally matched (in terms of CA, *IQ*, and length of institutionalization) groups of males and females.

Each *S* was seen individually, by the same examiner. During the course of a "routine" psychological evaluation, each *S* was asked to draw a Monster and to draw God (on blank sheets of 8½" × 11" paper, employing a No. 2 pencil). To reduce any immediate transfer effects, the Draw-God and the Draw-a-Monster tests were separated in time by being alternately administered at the beginning and the end of the evaluative session, in a counterbalanced fashion.

Upon completion of the Draw-God test, the following questions were asked: (*a*) Is God old or young? (*b*) What is God's age—how old is God? (*c*) What is God doing? (*d*) What kind of expression or feeling is on God's face? The following questions were asked when the Draw-a-Monster tests were completed: (*a*) Is the Monster old or young? (*b*) What is the Monster's age—how old is the Monster? (*c*) What is the Monster doing? (*d*) Is the Monster a human or an animal? (*e*) What kind or type of Monster is it?

C. RESULTS AND DISCUSSION

Table 1 depicts the results obtained here, in terms of chi square analyses. There were no statistically significant differences between the adolescent *S*s and the adult *S*s, in terms of the sizes of their drawn Gods and Monsters. However, within the adolescent group (but not within the adult group), a highly statistically significant difference was obtained between the sizes of the drawn Gods and Monsters, with the adolescent *S*s drawing significantly smaller Gods than Monsters. The adult *S*s also demonstrated a (nonsignificant) tendency in this same direction (i.e., viewing the Monster as larger than God).

TABLE 1
CHI SQUARE VALUES, ADOLESCENTS *vs.* ADULTS

Variable	Chi square
Size of God	2.80
Size of Monster	2.80
Size of God *vs.* Monster, adolescents	21.60**
Size of God *vs.* Monster, adults	2.40
Benevolent aspect of God	.09
Punishing aspect of Monster	6.80*
Human *vs.* animal Monster	.72

* Significant at .01 level of confidence.
** Significant at $< .001$ level of confidence.

This perception of God as being small (and also as not being a "helping" God) may symbolize these retardates' own feelings of helplessness and rejection. This negation of the benevolent aspects of God might be hypothesized as the symbolic equivalence of a counter-defensive maneuver designed to negate the effects of separation from the parents, parents who may be perceived of as not interested in helping the *S*s. The negative and perhaps hostile influences of the institution are, however, brutally perceived. In some way, the institution is a punishing monster, perhaps similar to the "punishing" parents who have "deserted" them. In addition, the adolescent *S*s would appear to react more strongly to the (adverse) institutional setting than do institutionalized adult retardates. One might hypothesize that institutionalization may be a more traumatic experience for adolescents than for adults.

In contrasting the adolescent and adult groups as to their perceptual-cognitive view of the aggressiveness of the drawn Monster, the results indicate that the adolescents perceive the Monster as being statistically more hostile (e.g., "kills people and eats them") than do adults, who view the Monster as being only mildly hostile (e.g., "scares people").

Klein believes that the pre-Oedipal stage of development is represented by oral-incorporation fantasies of an insatiable type. In accordance with this theory, then, we would expect the Monsters drawn by adolescents to reflect a killing-eating dyad more often than those done by adults, since oral-incorporative fantasies, according to Klein, become sublimated in adulthood, thereby losing their overt cannibalistic drive. Our data on this point, significant at the .01 level of confidence, tend to support this viewpoint.

Although no statistically significant results were attained in terms of the type of Monster drawn (human or animal), there was a tendency for the adolescent *S*s to portray more animal-type than humanoid-type Monsters. For example, some adult *S*s drew a "wolf-man," while some adolescents drew a "dinosaur." This tendency is in accord with Kleinian theory, as feminine-masculine symbolic animals belong primarily to the pre-Oedipal cimmerian stage of development, with masculine symbolic animals being characteristic of adulthood.

D. SUMMARY

Thirty institutionalized mildly mentally retarded adolescents and 30 adults were each asked to draw a Monster and to draw God. Based upon Freud's views, the authors assumed that the institution would stand as an authority figure and that attitudes toward the institution would be displaced onto these God and Monster drawings.

The findings support Freud and Klein's views. The adolescents produced significantly smaller drawings of God than they did of Monsters. They viewed the institution as a punishing Monster, similar to a punishing or deserting parent. The adults, as a possible consequence of ongoing frustrations of institutionalized life, developed a defense of perceptual indifference, and they viewed Monsters as less aggressive and more humanoid than did the adolescents. The theoretical implications of these findings are discussed.

REFERENCES

1. FREUD, S. Collected Papers. New York: Basic Books, 1960.
2. KLEIN, M. Contributions to Psycho-Analysis, 1921-1945. London: Hogarth, 1948.
3. KLEIN, M., *et al.* Developments in Psycho-Analysis. London: Hogarth, 1952.
4. LEDERER, W. Dragons, delinquents and destiny. *Psychol. Issues,* 1964, **4**, Monog. 15.
5. REIK, T. Der eigene und der fremde Gott. Vienna, Austria: Internat. Psychoanal. Verlag, 1923.

Willowbrook State School
2760 Victory Boulevard
Staten Island, New York 10314

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 305-312.

THE MMPI MEASUREMENT OF PHENOTYPIC ANXIETY: I. EXPLORATORY STUDY*[1, 2]

The University of Alabama and the University of Alabama Medical School

RAYMOND D. FOWLER, JR., F. A. COYLE, JR., P. C. REED, AND CHESTER A. WHITE

A. INTRODUCTION

Attempts to measure anxiety with psychometric instruments appear confounded partially due to a lack of appreciation of different types of anxiety. It has been suggested that there is a more basic, chronic "trait" anxiety which might be termed genotypic. This is opposed to the transient, situational "state" of anxiety which might be considered phenotypic. Levitt expressed this distinction as follows: "When a psychologist says that a person is anxious, the statement may be interpreted in either of two ways. It may mean the individual is anxious at the moment, or it may mean that he is an anxious person. These two interpretations are quite different. The former refers to an immediate and probably ephemeral state, whereas the latter is a constant condition without a time limitation" (5, p. 13).

Neglecting to separate these two types appears to have led to confusion in the interpretation of the various MMPI scales used to measure anxiety. The most widely used measure of anxiety was not intended as a clinical device when it was developed. Janet A. Taylor's (7) Manifest Anxiety Scale (MAS) was derived from a group of 200 MMPI statements. Five judges were instructed to select items indicative of manifest anxiety. The result of this procedure was a final version of the scale, consisting of 50 items. While the MAS does have a demonstrated ability to discriminate between groups which logically should differ, it unfortunately correlates quite highly with Scale 7 (Pt) of the MMPI. In fact, Brackbill and Little (1) found correlations between the MAS and Scale 7 as high as .92 with the psychiatric sample. Five more samples of patients confirmed this fact with correlations between .91 and .93. With a student group, correlation between these two scales was .81. With an alcoholic

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 19, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This research was supported, in part, by funds provided by Roche Laboratories, Division of Hoffmann-La Roche Inc.

[2] The authors would like to thank G. F. Tangoren, M.D., T. P. Springer and J. Johnston for their assistance in gathering the test data.

population (3), a correlation was found between the two scales of .92. In another study of college students, Deese, Lazarus, and Keenan (2) found a correlation of .81. Correlations of these magnitudes raise the question of whether the MAS contributes information in addition to that already provided by Scale 7. The Taylor MAS appears to be primarily based on the subject's willingness to admit the various types of physiological discomforts from which anxiety is referred. The general assumption is that overt anxiety is being assessed. In this case, there is no distinction made between the genotypic (chronic or basic) anxiety and phenotypic (situational or momentary) anxiety. In terms of behavioral consequences and clinical interpretation, this distinction seems to be a crucial one.

B. PROBLEM

In order to develop an MMPI scale which would differentiate between phenotypic and genotypic anxiety, an attempt was made to select a population that was being subjected to naturally occurring events which might be assumed to produce situational anxiety. It has been estimated (6) that surgery is performed on 12 million people in the United States each year. An earlier investigation (4) based on self-reports by patients found that more than 75 per cent reported a moderate to high degree of anxiety prior to the operation. The level of reported anxiety appeared to be related also to postoperative behavior. Patients reporting only a moderate amount of anxiety before developed fewer emotional symptoms following the operation.

C. METHOD

In view of the above, the first experiment was conducted with a group of 79 elective surgery patients at the University of Alabama Hospital. The subjects consisted of 62 female and 17 male preoperative patients in a diversified hospital environment, all of whom met the following criteria: age between 20 and 60; in good general health; no severe or unusual risks included in the operation, and no head or neck procedures involved. The MMPI was administered to each of these patients in two different sessions. The first administration occurred the night before the operation was to take place, during the time interval between the preparation of the patient and the administration of nightly medication.

The second administration occurred after the operation had been performed and the patient had recovered to the point where he had been taken off narcotics. Attempts were made to conduct the second testing the last night the subject was in the hospital.

Prior to the preoperative testing, a signed consent was obtained from each subject by the physician assisting in this experiment and an introductory explanation of the testing procedure was presented, assuring the patient of complete anonymity and that the results of the experiment were entirely irrelevant to the patient's medical and surgical care. The introduction contained no details concerning the objectives of the experiment.

The actual testing was initiated after visiting hours and following the preoperative preparation. Standard instructions were read to the patient who was then left to complete the forms.

The postoperative testing procedure was similar to the preoperative procedure with the exception of the elimination of the physician's assistance and subsequent consent form. Mean time between preoperative and postoperative testing was a four-day period and the mean time for the completion of the MMPI was 2.6 hours.

D. Results and Discussion

The profiles of both the pre- and postoperative tests were examined for each patient. If any of the validity scales were elevated greater than a T score of 80, both sets of data for that patient were removed. Some obvious cases of random responding were detected and those were also removed. The attrition rate was quite high and the final total number of subjects was 19.

As can be seen from Table 1, there were no appreciable differences in pre- and postoperative test scores for the patients on either the basic validity, clinical, or special scales. The MAS does not differentiate between the two test periods.

In an effort to develop a preliminary measure of phenotypic anxiety, a large sample of items was selected for inclusion in the preliminary scale for the measurement of phenotypic anxiety. An individual analysis of each item for the patients in each group was undertaken. This was believed desirable because of the limited number of the heterogeneous subjects which remained after the sizable rate of attrition. Because this is an exploratory study with extensive cross validation and refinement of the scale planned, the original scale is expected to undergo considerable modification.

Those items which differed by more than 10 per cent on the two testings were examined. If the number of subjects responding differently on the two testings was three or more, the item was retained, provided the response differentially separated performance on the first test from the second. In some cases the advantage is not clear-cut, but it is reasoned that each individual item can contribute an increment of difference. Thus, the sum total of responses on the items may result in a separation of the individuals on the scale. The items comprising the preliminary measure of phenotypic anxiety are listed in Table 2.

The scale was then scored for each group on the pre- and postoperative tests to see if it did differentiate. The t test performed yielded a $p > .001$. The scale does differentiate between the pre- and postoperative groups.

TABLE 1
PRE- AND POSTOPERATIVE MMPI SCALE SCORES
($N = 19$)

	Testing #1		Testing #2	
Scale	Mean	Standard deviation	Mean	Standard deviation
?	42.84	4.57	42.58	2.55
L	47.68	5.85	48.47	6.66
F	52.95	5.33	53.00	6.69
K	53.15	6.79	54.47	9.11
Hs	57.89	13.15	57.74	11.45
D	56.58	11.28	57.00	10.09
Hy	58.05	11.39	57.53	11.78
Pd	60.16	9.41	58.74	7.39
Mf	49.89	12.79	48.79	10.74
Pa	54.84	11.88	55.37	12.41
Pt	57.31	10.63	56.21	8.62
Sc	54.53	9.41	55.37	7.63
Ma	54.26	13.88	55.16	10.48
Si	56.11	9.01	55.21	8.57
Es	48.63	10.20	50.26	9.65
A	50.37	9.78	48.26	9.47
R	47.84	8.15	48.26	6.95
Lb	50.74	9.57	51.47	10.55
Ca	54.11	9.27	52.84	10.50
Dy	52.89	8.43	50.95	8.98
Do	50.05	8.67	53.53	9.51
Re	52.53	7.49	53.26	8.01
Pr	48.26	6.67	48.58	8.55
St	54.74	6.94	56.00	7.36
Cn	47.84	11.30	46.21	10.35
MAS[a]	20.32	9.53	18.16	9.36

[a] Raw score.

During the actual conduct of the experiment, there were only two major problem areas. The first was due to an insufficient screening of patients with regard to intelligence. In the future, it has been planned to control for this and eliminate, at the beginning, those patients obviously incapable of completing the test. The first statement on the MMPI is "I like Mechanics Magazine." Several patients misread or misinterpreted this to mean "I like machines." This could perhaps be due to the high level of anxiety, although

TABLE 2
THE 36 ITEMS OF THE PRELIMINARY MEASURE OF PHENOTYPIC ANXIETY, SHOWING THE DIRECTION OF ANSWERING WHICH IS TO BE SCORED

TRUE

4. "I think I would like to work as a librarian."
46. "My judgment is better than it ever was."
66. "I see things or animals or people around me that others do not see."
67. "I wish I could be as happy as others seem to be."
71. "I think a great many people exaggerate their misfortunes in order to gain the sympathy and help of others."
84. "These days I find it hard not to give up hope of amounting to something."
134. "At times, my thoughts have raced ahead faster than I could speak them."
166. "I'm afraid when I look down from a high place."
213. "In walking, I'm very careful to step over sidewalk cracks."
216. "There is very little love and companionship in my family as compared to other homes."
230. "I hardly ever notice my heart pounding and I'm seldom short of breath."
243. "I have few or no pains."
248. "Sometimes without any reason, or even when things are going wrong, I feel excitedly happy, 'on top of the world.'"
263. "I sweat very easily, even on cool days."
267. "When in a group of people, I have trouble thinking of the right things to talk about."
278. "I have often felt that strangers were looking at me critically."
296. "I have periods in which I feel unusually cheerful without any special reason."
397. "I have sometimes felt the difficulties were piling up so high that I could not overcome them."
398. "I often think, 'I wish I were a child again.'"
439. "It makes me nervous to have to wait."
489. "I feel sympathetic toward people who tend to hang on to their grief and trouble."
558. "A large number of people are guilty of bad sexual conduct."

FALSE

37. "I have never been in trouble because of my sex behavior."
40. "Most any time I would rather sit and daydream than to do anything else."
55. "I am almost never bothered by pain over the heart or in my chest."
83. "Any man who is able and willing to work hard has a good chance of succeeding."
122. "I seem to be about as capable and smart as most others around me."
159. "I cannot understand what I read as well as I used to."
172. "I frequently have to fight against showing that I'm bashful."
222. "It is not hard for me to ask help from my friends even though I cannot return the favor."
364. "People say insulting and vulgar things about me."
373. "I feel sure that there is only one true religion."
389. "My plans have frequently seemed so full of difficulties that I have had to give them up."
410. "I would certainly enjoy beating a crook at his own game."
449. "I enjoy gatherings just to be with people."
540. "My face has never been paralyzed."

in most instances an objective reason would be the lack of education or below average intelligence and verbal skills.

The second most disrupting problem was the apparent lack of communication, in some cases, between physician and patient after the initial contact concerning the MMPI. One female subject signed a consent form the morning prior to the scheduled evening test but evidently did not completely understand the complicated form. When the psychological worker appeared for the test session, the subject was so severely distraught that she became hysterical and was calmed only after the diligent effort of several doctors. Some patients had the mistaken idea that the type of surgery they were to undergo would be directly dependent upon the results of the MMPI. One patient nearly refused to submit because he was positive the test would increase his medical expenses. Emphasis must be placed upon establishing rapport and understanding at the initial contact with the patient. It should be noted that these events, while quite disruptive, occurred rarely, and the general level of cooperation was better than anticipated.

Some minor problems included visitors, but this was alleviated when the visiting hours were adjusted. Before this adjustment, the workers were sometimes forced to intrude upon the patient, and the patient's relatives resented the preemption of their visiting privileges. Interruptions by the medical staff were at a minimum. In most instances, the staff respected the testing procedure and interrupted only when necessary.

The results presented in the study are preliminary and experimental. They are by no means meant to be considered a final, valid indication of phenotypic anxiety. Additional investigation regarding both the meaningfulness and validity of such a scale is currently underway. Naturalistic settings in which there are realistic environmental stresses might be employed. Situations, such as a husband waiting for his wife to give birth, patients waiting to see the dentist, or parents whose children are about to undergo surgery, are being considered.

E. SUMMARY

Attempts to measure anxiety with psychometric instruments appear confounded, partially due to a lack of appreciation of different types of anxiety. It has been suggested that there is a more basic, chronic, "trait" anxiety which might be termed genotypic. This is opposed to the transient, situational, "state" form of anxiety which might be considered phenotypic. Neglecting to separate these two types appears to have led to confusion in the interpretation of the various MMPI scales used to measure anxiety. In order to develop an MMPI

scale which would differentiate between phenotypic and genotypic anxiety, an attempt was made to select a population which was being subjected to naturally occurring events which might be assumed to produce situational anxiety. The subjects consisted of 62 female and 17 male preoperative patients in a diversified hospital environment. The MMPI was administered to each of these patients in two different sessions. The first administration occurred the night before the operation was to take place, during the time interval between the preparation of the patient and the administration of nightly medication. The second administration occurred after the operation had been performed and the patient had recovered to the point that he had been taken off narcotics. There were no appreciable differences in pre- and posttest scores for the patients on either the basic validity, clinical, or special scales. An item analysis produced 36 items which appeared to differentiate between the two test periods. These items were assumed to constitute a preliminary scale for the identification of phenotypic anxiety. The results presented in the study are preliminary and experimental. They are by no means meant to be considered a final, valid indication of phenotypic anxiety. Additional investigation regarding both the meaningfulness and validity of such a scale is currently underway.

REFERENCES

1. BRACKBILL, G., & LITTLE, K. B. MMPI correlates of the Taylor scale of manifest anxiety. *J. Consult. Psychol.,* 1954, **18**, 433-436.
2. DEESE, J., LAZARUS, R. S., & KENNAN, J. Anxiety, anxiety reduction, and stress in learning. *J. Exper. Psychol.,* 1953, **46**, 55-60.
3. ENDS, E. J., & PAGE, C. W. Functional relationships among measures of anxiety, ego strength and adjustment. *J. Clin. Psychol.,* 1957, **13**, 148-150.
4. JANIS, I. L. Psychological Stress. New York: Wiley, 1958.
5. LEVITT, E. E. The Psychology of Anxiety. New York: Bobbs-Merrill, 1967.
6. ROTHENBERG, R. E. Understanding Surgery. New York: Pocket Books, 1966.
7. TAYLOR, J. A. A personality scale of manifest anxiety. *J. Abn. & Soc. Psychol.,* 1953, **48**, 285-290.

Department of Psychology
University of Alabama
University, Alabama 35486

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 313-319.

TRANSFER OF SOCIAL INFLUENCE EFFECTS ON PSYCHOPHYSICAL JUDGMENTS*[1]

Massachusetts Institute of Technology

ROBERT K. YIN AND HERBERT D. SALTZSTEIN

A. PURPOSE

Social influence procedures usually involve two elements: (*a*) stimuli about which a subject must make some judgments, and (*b*) exposure of the subject to discrepant judgments made by another person or persons. If the subject's own judgments about the stimuli are changed as a result of the exposure, then social influence is said to have taken place.

Most research on social influence has concentrated on demonstrating that changes in judgments of specific stimuli have indeed occurred (1, 5). Those few studies (2, 3, 4) that have been concerned with transfer effects have not specified the form of the transfer. If, however, the demonstrated social influence effects should transfer to new stimuli of the same class, then there must have been a change in the overall stimulus-judgment relationship for that entire class of stimuli.

The stimuli of the present study were chosen to take advantage of the psychophysical power relationship, which describes an orderly correlation between judgments of the intensities of stimuli and the actual physical magnitude of those stimuli. Investigators in psychophysics have found that a power function fits the data for a wide variety of continua, involving many different sensory modalities, and that each modality yields a characteristic exponent for the power function (8). The scaling of subjective roughness was chosen as the psychophysical task both because of its simplicity of administration and because it could be easily adapted to a social influence situation. A previous study (9) had shown that, with the use of various grades of emery cloth, the characteristic power function approximated 1.5.

Our question was whether subjects would actually exhibit a changed power

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 22, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

[1] This work was supported by grant number HD01786-01A1 from the National Institute of Child Health and Human Development to the second author, and from the National Aeronautics and Space Administration NsG496 to Dr. H.-L. Teuber. Drs. Rhea M. Diamond and Whitman Richards have read earlier versions of this manuscript and made valuable criticisms and suggestions.

function for the stimuli, after exposure to judgments representing a discrepant power function. This exposure was without overt reference to the relationships involved or the particular value of the exponent.

In order to demonstrate that the social influence effect had transferred in such a way as to result in a new power function, two conditions were imposed. First, the trials on which the subject was exposed to another person's judgment represented only a subset (and one that was of smaller range) of the original array of stimuli. These influence judgments came from a confederate who had been trained to give responses reflecting a power function different from 1.5. In all other respects, however, there was a minimum of social interaction between the subject and the confederate. Special attention was given to subsequent changes in the subject's judgments of stimuli *not used* in the influence situation.

Second, the only change examined was that shown during judgments made in private. We were not interested in the subject's performance in the presence of the confederate, but only in the difference between the subject's *preinfluence* power function (assessed privately in his first session and based on the whole set of stimuli) and his *postinfluence* power function (based on the same set and also assessed privately).

B. METHOD

1. *Subjects*

*S*s were 19 volunteer undergraduates, paid for their participation. Each *S* did the task in two sessions, set about one week apart.

2. *Task and Apparatus*

Stimuli used were 14 grades of aluminum oxide abrasive (Three-M-Ite Eleck-Tro-Cut Cloth, J Wt.) with the following mesh numbers: 36 (the roughest), 40, 60, 80, 100, 120, 150, 180, 220, 280, 320, 400, 500, and 600. The method used was cross-modality matching;[2] *S*s were asked to produce lines reflecting the apparent roughness of the abrasives (a longer line representing greater roughness). Cross-modality matching was chosen instead of the more usual magnitude estimation to reduce the likelihood that *S*s would remember specific estimates, and particularly to reduce the probability that *S*s might calculate the relationship between any of these estimates. The line was produced by moving one of two cords which were three inches apart and in parallel, circumscribing the girth of a two-foot wide box. (*E* was at the back of the box, recording the actual movement of the cord.) Stimuli were presented on the top of the box, in such a way that *S*s would feel the abrasives but not see them.

---

[2] In this method the subject matches sensations in one modality (with an unknown power function) to those of another (with a known power function). From his matches the unknown power function can be calculated. The method has been used extensively in psychophysics, and has yielded power functions similar to those derived by the method of magnitude estimation [see Stevens (7)].

3. *Procedure*

For the first session, *S*s were led to believe that the task was normally done in pairs (one cord for each *S*), but that the other member of the pair had not shown up or was late and would not be waited for. Each *S*'s first session was thus done individually, under the normal conditions of obtaining cross-modality matchings, with the 14 stimuli presented twice in two random orders, and without a modulus assigned by *E*.

For the second session, a confederate was present as the other member of the pair. Judgments were given in alternating turns for each trial (on one trial *S* would go first, followed by the confederate; on the next the confederate would go first, followed by *S*, and so on), with *S* always going first on the first trial so that he could assign his modulus without influence.

After seven of the original 14 stimuli (mesh numbers 40, 80, 120, 150, 220, 320, and 400) had been presented twice, the confederate stated that he had forgotten to make an important phone call and was allowed to leave the room on the condition that he would return immediately afterward to finish the experiment. *S* was told that there was no need for him to be delayed by waiting for the "other person" to return, so the entire array of 14 stimuli was now presented twice in random order to *S* alone. At the end of this series *S* was interviewed and fully apprised of the nature of the experiment.

4. *Experimental Groups*

After completion of the first sessions, the 19 *S*s were randomly assigned to three subgroups, two experimental and one control. Since the preinfluence means of these three subgroups were slightly different (although nonsignificantly), the subgroup with the highest mean was designated as Experimentals Up, to be influenced toward an even higher exponent for the power function, if possible; the one with the lowest mean was designated as Experimentals Down, to be influenced lower, if possible; and the middle subgroup became the control group.[3] This manner of designating the groups was followed in order to avoid any spurious effects due to regression toward the middle.

---

[3] Although each group originally had seven *S*s, two *S*s of the control group failed to return for the second session. This loss also changed the control group's mean preinfluence power function from 1.2 to 1.1, so that, contrary to our original plans, the control group did not turn out to be the group with the intermediate preinfluence mean.

A traditional control group (no exposure to a confederate's judgments) was not employed. Rather, our control subjects made their judgments in the presence of a confederate who provided judgments similar to their own. This permitted holding constant any general effects resulting from the presence of another person also making judgments, as opposed to the effects of the actual content of his responses.

The discrepant power function given by the confederate was not geared to the power functions of each individual *S*, but rather remained the same for all *S*s within a given subgroup. Thus, the confederate was trained to give responses approximating a power function of 2.1 in the presence of *S*s assigned to Experimentals Up, .3 in the presence of *S*s in Experimentals Down, and 1.2 in the presence of *S*s of the control group [1.2 representing the mean preinfluence power function of the original control group (see footnote 3)].

C. RESULTS

Table 1 shows the exponents of the power function for all *S*s in the three subgroups, pre- and postinfluence, and the differences between these two sets

TABLE 1
SUMMARY OF INDIVIDUAL POWER-FUNCTION EXPONENTS

Variable		Preinfluence exponent	Postinfluence exponent	Amount of change
Experimentals Up		.733	1.428	+.695
		1.241	1.515	+.274
		1.277	1.813	+.536
($N = 7$)		1.293	1.587	+.294
		1.371	1.844	+.473
		1.506	1.532	+.026
		1.691	2.024	+.333
	$\overline{X} =$	1.301	1.677	+.376
Experimentals Down		.688	.760	+.072
		.730	.749	+.019
		1.104	.568	—.536
($N = 7$)		1.267	1.041	—.226
		1.307	1.157	—.150
		1.444	1.114	—.330
		1.445	1.094	—.351
	$\overline{X} =$	1.141	.926	—.215
Control group		.691	1.158	+.467
		.927	.990	+.063
($N = 5$)		1.183	.950	—.233
		1.413	1.125	—.288
		1.447	1.316	—.131
	$\overline{X} =$	1.132	1.108	—.024

of exponents. The functions were derived by plotting the judgments of each *S* against the mesh size of each abrasive with both entries converted into log scores. The power function was then calculated as the slope of the straight line fit to these scores by the least squares method. With a *t*-test for dependent samples, matched pairs, it was found that $t = 4.310$ ($p < .005$, one-tailed) for Experimentals Up, $t = —2.460$ ($p < .025$, one-tailed) for Experimentals Down, and $t = —.160$ (not significant) for the control group.

In addition, the data for each of the three subgroups were analyzed according to the usual psychophysical procedure of taking the geometric mean of the group data for each stimulus, and plotting the points on log-log coordinates. Figure 1 shows the three sets of graphs. The preinfluence exponents for Experimentals Up, Experimentals Down, and the control group were 1.31, 1.12, and 1.14, respectively, while the corresponding postinfluence functions were 1.67, .93, and 1.16.[4]

This second procedure enabled us to see whether the responses made to the particular stimuli previously used during the influence trials, abrasives 40, 80, 120, 150, 220, 320, and 400, formed a different postinfluence pattern from the rest of the stimuli. This was to check the possibility that changes in judgments on these stimuli alone were the main determinants of the new power function. As the graphs show, this interpretation is unsupported.

D. DISCUSSION

Although the data indicate that generalization of social influence has occurred, several questions remain about the precise nature of the transfer. For instance, the influence subset of stimuli was arbitrarily chosen, and it may be that influence on other subsets would lead to different kinds of transfer.

Another issue is the extent to which any transfer is task specific. It would be interesting to see whether social influence on one psychophysical task would have any effect on the performance of another task with a different sensory modality or a different judgmental modality. In addition, in the present experiment no attempt was made to assess the power functions of subjective roughness and line production independent of each other, and there is thus no way of determining what part of the total change in the postinfluence trials might be attributable to a change in subjective roughness and what part might be attributable to a change in *S*'s scale for line production.

---

[4] The slight discrepancy between all six of these power functions and the six mean scores shown on Table 1 reflects the two different ways of grouping the data, the former using the geometric means of all the judgments for each stimulus and then deriving the power function from the stimuli scores, and the latter (as in Table 1) deriving the mean power function from the power functions calculated for each individual subject.

Finally, a psychophysical task, while providing a relatively simple and straightforward measure of the transfer of social influence, is still somewhat artificial and arbitrary when we consider the problems an individual faces in adapting to his social world. If the transfer phenomenon studied here is to have meaning beyond the confines of the laboratory, then it must be shown that comparable changes can be effected in tasks involving psychosocial scales or other measures of judgments formed primarily in social contexts. We plan to pursue these and related questions further in our follow-up experiments.[5]

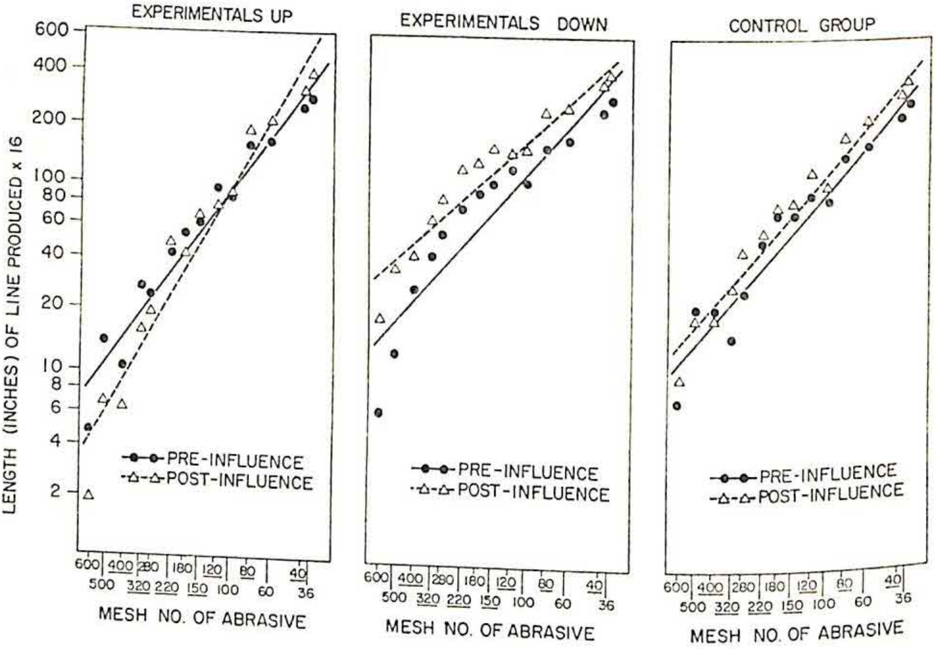


FIGURE 1
THE GEOMETRIC MEANS OF THE CROSS-MODALITY MATCHES TO SUBJECTIVE ROUGHNESS PLOTTED ON LOG-LOG COORDINATES

Mesh numbers of stimuli used in the influence trials between the pre- and post-influence series are underlined.

---

[5] The experimental results raise still another question, which is perhaps more relevant to the study of psychophysics than to the study of social influence. What would have happened if the confederate's judgments had not taken the form of a power function but had described some other mathematical function? Stevens seems to believe that the power function is a natural characteristic of the transducer which converts physical energies into psychological processes culminating in the observer's judgments [see Stevens (6, p. 28)]. This would suggest that no set of judgments by a confederate—no matter what curve they formed—could induce an observer to produce judgments which did not best fit a power function.

E. SUMMARY

Systematic changes in judgments on a psychophysical task were found to occur from one private session to another as a result of exposure to discrepant judgments of a confederate during an intervening session. The changes obtained included direct effects and transfer to new stimuli. The stimuli were different grades of aluminum oxide abrasives, presented to obtain cross-modality matchings of subjective roughness with length of lines. Each *S* made a series of such judgments in an initial private session, yielding an individual relationship typical of the psychophysical power function. In a second session, each *S* repeated the task twice. First, *S*'s judgments for a *subset* of the original stimuli were interspersed with those of a confederate trained to produce discrepant responses approximating a different power function. Second, after the confederate left, *S* made judgments of the entire original set of stimuli. Results were assessed in terms of the differences between each *S*'s pre- and postinfluence power functions. It was found that the power functions of one group of seven *S*s had changed upward after exposure to judgments discrepant in an upward direction ($p < .005$, one-tailed); another seven *S*s exposed to a downward influence treatment also showed an effect ($p < .025$, one-tailed), while those of a control group of five *S*s did not change.

REFERENCES

1. ASCH, S. E. Studies of independence and submission to group pressure: I. A minority of one against a unanimous majority. *Psychol. Monog.*, 1956, **70**(9), Whole No. 416.
2. FISHER, S., & LUBIN, A. Distance and influence in two person interaction. *J. Abn. & Soc. Psychol.*, 1958, **56**, 230-238.
3. PETERSON, D. E., SALTZSTEIN, H. D., & EBBE, C. Sequential effects in social influence. *J. Personal. & Soc. Psychol.*, 1967, **6**, 169-174.
4. SALTZSTEIN, H. D., ROWE, P. B., & GREENE, M. E. Spread of social influence on children's judgments of numerosity. *J. Personal & Soc. Psychol.*, 1966, **3**, 665-674.
5. SHERIF, M. The Psychology of Social Norms. New York: Harper & Bros., 1936.
6. STEVENS, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory Communication*. Cambridge, Mass.: M.I.T. Press. Pp. 1-33.
7. ———. The surprising simplicity of sensory metrics. *Amer. Psychol.*, 1962, **17**, 29-39.
8. STEVENS, S. S., & GALANTER, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. Exper. Psychol.*, 1957, **54**, 377-411.
9. STEVENS, S. S., & HARRIS, J. R. The scaling of subjective roughness and smoothness. *J. Exper. Psychol.*, 1962, **64**, 489-494.

Department of Psychology
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Published as a separate and in *The Journal of Psychology*, 1968, **68**, 321-328.

JUDGMENTS OF ETHNIC GROUPS*

University of Kentucky

JOSEPH C. FINNEY

A. INTRODUCTION

In 1948, using an adjective checklist, Vinacke (7) showed that each of eight ethnic groups in Hawaii had a distinct image or stereotype.

The present investigator noticed that many of the adjectives or phrases found by Vinacke fell into groups that suggested psychoanalytic character types (3) and that psychiatrists and clinical psychologists living in Hawaii applied such terms to the ethnic groups: "The Japanese are compulsive characters." "The Hawaiians are very noncompulsive." "The Filipinos do a lot of hysterical acting out." The psychoanalytic terms used seemed consistent with the sets of adjectives found by Vinacke.

It was decided to test the hypothesis that the differences in the perceptions of the several ethnic groups in Hawaii could be understood largely in terms of psychiatric character types or personality types. It was considered likely that much of the image of the Japanese, for example, consisted of people perceiving, with some validity and truth, that Japanese tend to be high on the dimension that psychiatrists call compulsive character. The subjects would probably not be aware that they were perceiving in terms of such a dimension, and would probably not have a name for the dimension, nor would they probably be aware that such traits as orderliness, thrift, and stubbornness (the triad classically described by Freud) share a common loading on that dimension.

It was also desired to replicate the study of Vinacke and to see what changes, if any, had occurred in the 15-year interval between the former study and the present one carried out in 1962 and 1963.

The following predictions were made:

1. The descriptions of other ethnic groups will be consistent with previously reported opinions, including those found in Vinacke's study.

2. Some of the major differences in the groups, as described, will reflect psychoanalytic character types. Some groups (Filipinos, Puerto Ricans, Portuguese) will be described as hysterical, phallic, dissociating characters, while

---

* Received in the Editorial Office, Provincetown, Massachusetts, on January 23, 1968, and published immediately at 35 New Street, Worcester, Massachusetts. Copyright by The Journal Press.

others (Chinese, Japanese, Koreans) will be described as compulsive, anal, reaction-forming characters.

3. Insofar as a separation of content from desirability can be made, the ingroup and the outgroup may agree in the content of describing that group, while differing in the judgment of its desirability.

B. Method

A group of 257 students taking an introductory course in psychology at the University of Hawaii served as subjects by rating 12 objects on 50 contrasting pairs of phrases, using a seven-point scale—essentially the procedure that Osgood (6) calls the "semantic differential." The objects rated were oneself as an individual and 11 ethnic groups living in Hawaii: Portuguese, other Caucasian ("haole"), Chinese, Negro, Filipino, Hawaiian, Hawaiian-haole, Hawaiian-Chinese, Japanese, Korean, and Puerto Rican. The subjects were 72 men and 185 women: 135 Japanese, 38 Chinese, 29 haoles, 24 part-Hawaiians, nine Koreans, six Filipinos, five Portuguese, and 11 unknown or unclassifiable.

Table A[1] shows the 50 pairs of contrasting phrases. Some of the phrases were included because they were heard in popular usage to describe ethnic groups in Hawaii, or were found to do so in Vinacke's study. Other phrases were composed to define psychiatric types. For example, two were designed to define the paranoid character whose chief defense is projection: *feels mistreated* and *resentful*. Some 22 phrases were considered to define the compulsive or anal character whose chief defense is reaction formation: among them were Freud's (5) triad of *orderly, systematic; thrifty;* and *stubborn, rigid;* as well as *clean, neat; tense, worries; consistently works hard; plans years ahead;* and *self-reliant*. The seven items defining the hysterical or phallic character whose chief defense is dissociation-repression-denial were *impulsive, naive, shows off (flashy if extreme), dramatic, acts intuitively, flirtatious,* and *hot-tempered*. By hindsight, the phrases defining hysterical character seem too heavily loaded with assertiveness, forming a combination known as hysterical acting out. If the study were repeated, some phrases describing quieter forms of the hysterical character would be included, such as *romantic, youthful, childlike, idealistic, dreamer, artistic,* and *having faith in people (not skeptical)*.

---

[1] Tables A, B, and C have been deposited as Document number 9817 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington, D.C. 20540. A copy may be secured by citing the Document number and by remitting $1.25 for photoprints, or $1.25 for 35-mm microfilm. Advance payment is required. Make checks or money orders payable to Chief, Photoduplication Service, Library of Congress.

The terms that were composed or selected to represent the psychodynamic character traits had to be terms that a psychoanalytically oriented psychiatrist or clinical psychologist would recognize as descriptive of a character type or dimension, such as compulsive personality or hysterical personality; and at the same time they had to be nontechnical terms that an untrained person would feel able to use in describing people.

By combining various items, scales of compliance *versus* assertion (25 items), conscientiousness (14 items), hostility (nine items), desirability (17 items), activity (18 items), and strength (12 items) were established.

In the analysis, the ratings of each group have been separated into ratings by members of the group (ingroup ratings) and ratings by all persons not members of the group (outgroup ratings). Critical ratios have been calculated for the comparison of the outgroup ratings given to each group with the combined outgroup ratings given to all other groups on the same trait-name. Two matrices have been drawn up (Tables B and C),[2] one composed of the means of the ingroup ratings and the other of the means of the outgroup ratings. Each matrix has 50 rows for the traits and nine or 11 columns for the object groups. The phrase "within rows," elsewhere in this paper, refers to those two matrices.

The image of each group correlated with the image of every other group, with the mean score of each group for each of the 50 qualities, as perceived by its pooled outgroups.

C. Results

Table 1 shows the intercorrelations of the images of the 11 ethnic groups. There were two clusters of three groups each, which correlated highly positively within each cluster and negatively between one cluster and the other.

TABLE 1

INTERCORRELATIONS OF THE IMAGES OF THE 11 ETHNIC GROUPS

	Jap.	Kor.	Chin.	hao.	H.-C.	H.-h.	Haw.	Neg.	Fil.	P.R.
Kor.	+.94									
Chin.	+.81	+.73								
hao.	+.02	—.12	+.22							
H.-C.	+.13	+.20	—.15	+.05						
H.-h.	—.15	—.10	—.41	+.08	+.79					
Haw.	—.41	—.32	—.74	—.27	+.62	+.63				
Neg.	—.46	—.41	—.52	—.51	—.28	—.15	+.26			
Fil.	—.73	—.68	—.60	—.42	—.50	—.24	+.10	+.49		
P.R.	—.83	—.79	—.64	—.31	—.52	—.22	+.08	+.55	+.94	
Port.	—.87	—.85	—.63	—.08	—.41	—.13	+.12	+.33	+.78	+.90

---

[2] See footnote 1.

One was composed of Japanese, Chinese, and Koreans, and the other of Portuguese, Puerto Ricans, and Filipinos. A third cluster was composed of the three Hawaiian groups.

With few exceptions, the phrases found in Vinacke's study still characterized the images of the groups. The most striking exception was "hot-tempered," which was no longer part of the ingroup and outgroup image of the Koreans.

Total scores for each group on the 22 reaction-forming (or compulsive) qualities were as follows (the neutral rating is 110): Chinese, 203; Japanese, 175; Koreans, 159; haoles, 136; Negroes, 105; Hawaiian-Chinese, 100; Hawaiian-haoles, 95; Filipinos, 83; Puerto Ricans, 66; Portuguese, 63; Hawaiians, 25. All of the predictions about the reaction-forming or compulsive characters were confirmed.

Total scores for each group on the seven dissociating (or hysterical) qualities were as follows (the neutral rating is 35): Portuguese, 60; Filipinos, 59; Puerto Ricans, 51; haoles, 49; Hawaiians, 36; Negroes, 35; Hawaiian-haoles, 35; Hawaiian-Chinese, 21; Chinese, 15; Koreans, 12; Japanese, 12.

A measure of compulsivity *versus* hysteria, based on rank orders, yielded the following scores (neutral is 10): Japanese, 18½; Chinese, 18; Koreans, 17½; Hawaiian-Chinese, 12; Negroes, 12; Hawaiian-haoles, 10; haoles, 10; Hawaiians, 4; Filipinos, 4; Puerto Ricans, 4; Portuguese, 1. A similar measure of combined use of the two defense mechanisms showed that haoles were seen as most defended and Hawaiians least, as follows: haoles, 14; Chinese, 12; Filipinos, 12; Portuguese, 11; Negroes, 11; Puerto Ricans, 10; Japanese, 9; Koreans, 9; Hawaiian-Chinese, 8; Hawaiian-haoles, 8; and Hawaiians, 6.

Similar computation on the ratings of one's own group yielded, for compulsive character (neutral rating 88): Chinese, 150; Japanese, 135; Koreans, 106; haoles, 95; Filipinos, 89; Hawaiian-Chinese, 72½; Portuguese, 63½; Hawaiian-haoles, 58; Hawaiians, 23.

Similar computation on the ratings of one's own group yielded, for hysterical character (neutral rating 28): Portuguese, 49; Filipinos, 42; haoles, 41; Hawaiians, 29; Hawaiian-haoles, 27; Koreans, 23; Hawaiian-Chinese, 22½; Japanese, 17; Chinese, 1½.

D. DISCUSSION

1. *Psychodynamic Character Types*

The prediction was confirmed that scales composed of items designed to tap descriptions of personality dimensions, such as compulsive (anal, reaction-forming) character or hysterical (phallic, dissociating) character, would distinguish the images of the several groups. The predicted differences were found, and they characterized the expected groups in the expected directions. The three "Oriental" groups—Chinese, Japanese, and Koreans—were rated as compulsive personalities, and their images correlated high with each other's. Likewise, three other groups—the Filipinos, the Puerto Ricans, and the Portuguese—were rated as hysterical personalities, and their images correlated highly with each other's. The Hawaiians and the haoles also come out as predicted on those dimensions.

There was a general tendency for groups that scored high on compulsive character to score low on hysterical character ($r = —.65$ for outgroup ratings). This trend was most marked for the three Oriental groups (Chinese, Japanese, Koreans), who scored high on compulsiveness and low on hysteria. On the other hand, the haoles scored above the mean on both traits, while the Hawaiians scored extremely low on compulsiveness and only average on hysteria.

On both compulsive and hysterical characters, the rank order of groups, when each ethnic group was rated by its own members, was almost identical with the rank order obtained from ratings of each ethnic group by outsiders. Groups that were considered compulsive by outsiders were also considered compulsive by their own members, and the same principle held true for hysterical character.

The prediction that the Chinese, Japanese, and Koreans would be perceived as similar to each other was a natural one. Though people in Hawaii speak of some differences among them, the three groups are regarded as related both in race and in culture, and are commonly lumped together as "the Orientals," a term that does not include the Filipinos.

But the prediction that the Filipinos, the Puerto Ricans, and the Portuguese would be perceived as similar to each other was much less obvious. There is no common term that lumps them together. Racially they are quite different. Though all three groups are believed to be racially mixed, the Filipinos are chiefly of Mongoloid race, the Portuguese chiefly Caucasian, and the Puerto Ricans are largely, if not chiefly, of Negro ancestry. The social status of the three groups is different, too: the Portuguese rank highest of the three, and the Puerto Ricans lowest. The prediction of the present investigator that the three groups would be perceived alike was based on surmising that the popular picture of all three groups taken separately seemed to be accounts of behavior that could be described psychiatrically as hysterical personality (or dissociated personality).

The most obvious thing that the Portuguese, the Filipinos, and the Puerto

Ricans have in common is their Spanish (or at least Iberian) background. The hypothesis suggests itself that their common Spanish culture may be the source of their common hysterical character, as perceived by others and by themselves.

2. *Value, Fact, and Projection in Judgment*

The question arises whether the perceptions about the groups, the images held about the groups, bear a substantial relationship to the truth. It has been customary to stress the element of consistent nonrandom error known as racial prejudice, and to stigmatize all images as "stereotypes" with the implication that they are both false and pejorative. Perhaps this is because so many studies have dealt with white-Negro and Gentile-Jewish judgments, highly laden with emotion, and, probably, with projection. In Hawaii, however, the antagonisms among the several groups are in low key, and there is good reason to believe that the differences among images of the groups reflect a fairly accurate perception of differences in culture or ways of life.

The agreement of ingroup judgments with the outgroup judgments is not a sure sign that the judgment is true, however. The members of a group may be influenced by outsiders into accepting a false view of their own group or of themselves (cf., Cooley's "looking-glass self"). Even if the judgment is shown to be true, it is not necessarily a simple perception of the facts about one's own culture or about oneself. People's behavior may conform to what they have learned is expected of their group. The relationship between image and behavior is a circular one.

For low-key situations the hypothesis is advanced that each judgment is the resultant of two orthogonal factors: one representing perception of fact; and the other representing desirability judgment, Osgood's (6) "evaluative" factor of good-*versus*-bad. Whenever that is the case, one can measure the value judgment, remove it mathematically, and have some factual observation remaining. When the situation requires the correction of a score, rather than a correlation coefficient, a correction formula of Finney (4) can be used, for scores in standard form: $T = S - r\,A/\sqrt{1 - r^2}$.

It may well be that two adjectives may have the same objective content, may describe the same behavior, and yet may have opposite desirability values, one good and the other bad. Such may be the case with *thrifty* and *stingy,* with *firm* and *stubborn,* with *consistent* and *rigid,* or with *flexible* and *unreliable.* The extent to which such terms are factually equivalent can be determined by using both in the same rating study, and then removing the desirability factor mathematically.

In terms of this schema, one can expect to find people in apparent disagreement who agree on the facts but disagree on the value judgment. A fresh look at some of Vinacke's data confirms this hypothesis. Japanese and haoles, as well as Hawaiians, agreed in calling Hawaiians "easy-going" and "happy-go-lucky," but haoles and Hawaiians regarded these terms favorably, while Japanese did not. Furthermore, the Japanese called the Hawaiians "lazy," while the haoles did not. Everyone regarded the term "lazy" as unfavorable. It seems clear that the terms "easy-going" and "lazy" had the same objective denotative content (described the same behavior), while differing in value judgment (favorable *versus* unfavorable). The several groups agreed in ascribing the same behavior to the Hawaiians, but disagreed in how favorably or unfavorably they regarded it [see Finney (2)].

The present study did not permit testing that hypothesis further. The concepts "lazy" and "easy-going" were combined into a single trait-name, "consistently works hard" *versus* "easy-going, happy-go-lucky," which did not have a strong evaluative factor. Both "stingy" and "thrifty" were included, but their equivalence in content was destroyed by pitting "stingy" against "generous" (implying spending on others), while pitting "thrifty" against "spends money freely" (implying spending on oneself).

While low-key judgments may be accounted for by two factors, objective fact and subjective desirability, the prejudicial stereotyping that is highly emotionally charged seems to involve a third factor, one of projection. This is a process by which one attributes to others the things that one does not like to perceive in himself. When projection occurs, it cannot be expected that objective judgments will be obtained from partialing the desirability factor out of the raw judgments. Indeed, the extent to which the mathematical correction fails may be a measure of the extent to which projection is occurring.

E. SUMMARY

In a study of how people of various ethnic groups judge each other, students in Hawaii were asked to rate 11 groups and themselves on 50 pairs of contrasting phrases on seven-point scales. Vinacke's earlier findings were confirmed and extended. A hypothesis that the perceived differences among groups corresponded to psychoanalytic character types was tested and confirmed. Three ethnic groups were perceived as alike and as compulsive or reaction-forming characters, while three other ethnic groups were perceived as alike and as hysterical or dissociating characters. It was suggested that judgments contain at least two or three elements: a factual element, an evaluative element, and sometimes also a projective element.

REFERENCES

1. COOLEY, C. H. Social Process. New York: Scribner, 1918.
2. FINNEY, J. C. Attitudes of others in Hawaii towards Hawaiians. *Soc. Process,* 1962, **5**, 78-83.
3. ———. Psychiatry and multiculturality in Hawaii. *Internat. J. Soc. Psychiat.,* 1963, **9**, 5-11.
4. ———. Correction for unwanted variance. Unpublished paper read at the American Psychological Association, New York, 1966.
5. FREUD, S. Character and anal erotism (1908). In *Collected Papers* (*Vol. 2*). London: Hogarth, 1950. Chap. 4.
6. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. The Measurement of Meaning. Urbana, Ill.: Univ. Illinois Press, 1957.
7. VINACKE, W. E. Stereotyping among national racial groups in Hawaii: A study in ethnocentrism. *J. Soc. Psychol.,* 1949, **30**, 265-291.

College of Education
University of Kentucky
Lexington, Kentucky 40506

REFERENCES

1. COOLEY, C. H. Social Process. New York: Scribner, 1918.
2. FINNEY, J. C. Attitudes of others in Hawaii towards Hawaiians. *Soc. Process*, 1962, **5**, 78-83.
3. ———. Psychiatry and multiculturality in Hawaii. *Internat. J. Soc. Psychiat.*, 1963, **9**, 5-11.
4. ———. Correction for unwanted variance. Unpublished paper read at the American Psychological Association, New York, 1966.
5. FREUD, S. Character and anal erotism (1908). In *Collected Papers* (*Vol. 2*). London: Hogarth, 1950. Chap. 4.
6. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. The Measurement of Meaning. Urbana, Ill.: Univ. Illinois Press, 1957.
7. VINACKE, W. E. Stereotyping among national racial groups in Hawaii: A study in ethnocentrism. *J. Soc. Psychol.*, 1949, **30**, 265-291.

College of Education
University of Kentucky
Lexington, Kentucky 40506